

**TOPICS IN LABOR AND EXPERIMENTAL ECONOMICS**

by

**Rachel B. Landsman**

Bachelor of Arts, Economics, CSU Sacramento, 2012

Master of Arts, Economics, University of Pittsburgh, 2014

Submitted to the Graduate Faculty of  
the Dietrich School of Arts and Sciences in partial fulfillment  
of the requirements for the degree of

**Doctor of Philosophy**

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH  
DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Rachel B. Landsman

It was defended on

June 26th 2018

and approved by

Lise Vesterlund, Department of Economics, University of Pittsburgh

Rania Gihleb, Department of Economics, University of Pittsburgh

David Huffman, Department of Economics, University of Pittsburgh

Sera Linardi, Graduate School of Public and International Affairs, University of Pittsburgh

Dissertation Advisors: Lise Vesterlund, Department of Economics, University of Pittsburgh,

Rania Gihleb, Department of Economics, University of Pittsburgh

## **TOPICS IN LABOR AND EXPERIMENTAL ECONOMICS**

Rachel B. Landsman, PhD

University of Pittsburgh, 2019

This dissertation consists of three essays on labor and experimental economics. The first two chapters address gender differences in labor markets. The third chapter is a methodological contribution to experimental economics. Chapter 1 presents work from an experimental study on gender differences in returns to negotiation. The study presents a first examination of a negotiation-ban counterfactual. The study looks at manager-selected compensation when workers can and cannot negotiate their salary in an environment where workers perform either high or low productivity tasks. We find that when negotiation is allowed, a gender earnings gap emerges among high productivity workers. A similar gender earnings gap is not present when negotiation is not permitted. Chapter 2 presents work from a study on gender differences in executive departure rates. The study asks whether the higher departure rates of female executives relative to male executives can be explained by differences in ability or if other factors are also contributing to the departure rate gap. Using exogenous changes in firm performance coming from industry wide fluctuations, the study demonstrates that the departure rate gap cannot be explained by ability alone. The study provides evidence that, instead, the gap is consistent with attribution bias. Chapter 3 presents work from a study that tests whether the slider task is responsive to incentives. Using a between-subject design with three different piece rates: half a cent, two cents, and 8 cents, we find that the performance response in the slider task is very inelastic to incentives. Following a 1500% increase in incentives, output only increased by 5%. This study cautions researchers that the slider task may be underpowered in between-subject experiments that use typical experimental sample sizes.

## TABLE OF CONTENTS

<b>PREFACE</b> . . . . .	ix
<b>1.0 BANNING NEGOTIATION: IS DIFFERENTIAL PAY ELIMINATED OR SUSTAINED WHEN LEFT TO MANAGER DISCRETION? (CO-AUTHORS: RANIA GIHLEB AND LISE VESTER-LUND)</b> . . . . .	1
1.1 Introduction . . . . .	1
1.2 Experimental Design . . . . .	3
1.2.1 Production Stage and Work Tasks . . . . .	4
1.2.2 Negotiation Stage . . . . .	5
1.2.3 Payment-Selection Stage . . . . .	6
1.2.4 Follow-up Activities . . . . .	6
1.3 Data . . . . .	6
1.4 Results . . . . .	7
1.4.1 How do Managers Compensate Workers? . . . . .	7
1.4.2 What are the Institutional Effects of a Negotiation-Ban? . . . . .	14
1.4.3 Why does a Negotiation-Ban Reduce Compensation Inequality? . . . . .	15
1.4.4 Gender . . . . .	17
1.4.4.1 Overview of Mixed-Gender Group Subsample . . . . .	20
1.4.4.2 Do the Institutional Effects of a Negotiation-Ban Differ by Gender? . . . . .	22
1.4.4.3 Why do the Effects of a Negotiation-Ban Differ by Gender Composition? . . . . .	24
1.5 Discussion . . . . .	26
1.5.1 Does Negotiation Affect Manager Beliefs? . . . . .	26
1.5.2 Differences in Chat Style . . . . .	28
1.5.3 Is the Gender Gap in Returns to Negotiation Attenuated by Differences in Negotiation Style? . . . . .	32
1.6 Conclusion . . . . .	36
<b>2.0 GENDER DIFFERENCES IN EXECUTIVE DEPARTURE</b> . . . . .	37

2.1	Introduction . . . . .	37
2.2	Related Literature . . . . .	39
2.2.1	The Executive Labor Market . . . . .	39
2.2.2	Executive Departure . . . . .	39
2.3	Data . . . . .	40
2.3.1	Data Overview . . . . .	40
2.3.2	Identifying Executive Departures . . . . .	41
2.3.3	Firm Performance . . . . .	43
2.3.4	Summary Statistics . . . . .	43
2.4	Methodology . . . . .	45
2.5	Results . . . . .	48
2.5.1	Main Result . . . . .	48
2.5.2	Robustness Checks . . . . .	51
2.6	Mechanisms . . . . .	55
2.6.1	Fertility . . . . .	55
2.6.2	Early Retirement . . . . .	56
2.6.3	Glass Cliff . . . . .	61
2.6.4	External Hires . . . . .	63
2.6.5	Female Start-Ups . . . . .	66
2.7	Misplaced Blame . . . . .	67
2.8	Conclusion . . . . .	68
3.0	<b>THE SLIDER TASK: AN EXAMPLE OF RESTRICTED INFERENCE ON INCENTIVE EFFECTS (CO-AUTHORS: FELIPE A. ARAUJO, ERIN CARBONE, LYNN CONELL-PRICE, MARLI W. DUNIETZ, ANIA JAROSZEWICZ, DIEGO LAMÉ, LISE VESTERLUND, STEPHANIE W. WANG, AND ALISTAIR J. WILSON)</b> . . . . .	69
3.1	Introduction . . . . .	69
3.2	Experimental Design . . . . .	71
3.3	Results . . . . .	72
3.4	Discussion . . . . .	77
3.5	Conclusion . . . . .	78
	<b>APPENDIX. ADDITIONAL FIGURES</b> . . . . .	80
	<b>BIBLIOGRAPHY</b> . . . . .	82

## LIST OF TABLES

1.1	Summary Statistics . . . . .	10
1.2	Gini Coefficient and 90-10 Ratio, By Treatment . . . . .	15
1.3	Per-Period Worker Earnings: Negotiation Treatment . . . . .	18
1.4	Worker Compensation Pooling over Both Treatments . . . . .	19
1.5	Summary Statistics: Mixed Gender Pairs . . . . .	21
1.6	Gini Coefficient and 90-10 Ratio, By Treatment and Gender Composition . . . . .	23
1.7	Per Period Worker Earnings: Negotiation Treatment, Mixed Gender Groups . . . . .	25
1.8	Typing Worker Compensation Pooling over Both treatments: Mixed Gender Pairs . . . . .	27
1.9	Stated Beliefs . . . . .	29
1.10	Effect of Manager Beliefs on Typing-Worker Pay . . . . .	30
1.11	Manager Beliefs on Typing-Worker Pay: Mixed-Sex . . . . .	31
1.12	Frequency of Characteristics Among Workers that Negotiated . . . . .	33
1.13	Frequency of Characteristics Among Workers that Negotiated: Mixed Gender . . . . .	34
1.14	Typing-Worker Pay Controlling for Chat Characteristics: Mixed Gender . . . . .	35
2.1	Summary Statistics . . . . .	44
2.2	Departure Probability: Results of Second-Stage Regressions . . . . .	50
2.3	Departure Probability with Continuous Performance Measures: Results of Second-Stage Regression . . . . .	52
2.4	Sensitivity of Departure Probability to Measure of Industry Shocks: Results of Second Stage Regressions . . . . .	53
2.5	Additional Robustness Tests: Results of Second-Stage Regressions . . . . .	54
2.6	Subsample Analysis using Older Executives: Results of Second Stage Regressions . . . . .	57
2.7	Subsample Analysis using Younger Executives: Results of Second Stage Regressions . . . . .	62
2.8	Subsample Analysis using Executives Hired when Firm is Performing Well Relative to Peers: Results of Second-Stage Regressions . . . . .	64
2.9	Subsample Analysis using Executives with More Tenure: Results of Second-Stage Regressions . . . . .	65
3.1	Summary Statistics . . . . .	72

3.2	Random-Effect Regressions . . . . .	75
3.3	Power Calculations . . . . .	76

## LIST OF FIGURES

1.1	Output over Time by Worker Task . . . . .	8
1.2	Output over Time by Worker Task and Treatment . . . . .	9
1.3	Distribution of Worker Output and Worker Earnings . . . . .	11
1.4	Earnings over Time by Worker Task . . . . .	12
1.5	Relative Production and Compensation over Time . . . . .	13
1.6	Share of Workers who Initiate Negotiation over Time by Worker Task . . . . .	16
2.1	Departure by Gender and Exogenous Performance . . . . .	48
2.2	Percentage of Female Entrants by Year . . . . .	58
2.3	Average Age of Entrants by Year and Gender . . . . .	59
2.4	Average Age of Executives by Year, Gender, and Departure Status . . . . .	60
3.1	Output Across Rounds . . . . .	73
A.1	Difficulty Adjusted Compensation - First Four Periods . . . . .	80
A.2	Difficulty Adjusted Compensation - All Five Periods . . . . .	81



## **PREFACE**

This dissertation would not have been possible without the immense support I received from my husband, Gabriel Routh, and my outstanding advisors, Lise Vesterlund and Rania Gihleb. I am also grateful for the frequent venting sessions, brainstorming sessions, feedback, and much needed breaks provided to me by my fellow graduate students, Mallory Avery, Jessica LaVoice, Sijie Li, Diego Lamé, Erin Carbone, Felipe Araujo, Neeraja Gupta, and many others. I am forever humbled by the love and encouragement I received from my personal community. I would also like to acknowledge the helpful feedback I received from my full dissertation committee, my coauthors, and countless seminar participants in the Economics Department at The University of Pittsburgh.

## 1.0 BANNING NEGOTIATION: IS DIFFERENTIAL PAY ELIMINATED OR SUSTAINED WHEN LEFT TO MANAGER DISCRETION? (CO-AUTHORS: RANIA GIHLEB AND LISE VESTERLUND)

### 1.1 INTRODUCTION

Does allowing for salary negotiation lead to increased wage inequality among workers? According to 84% of adults recently surveyed in the United States, the answer is yes.<sup>1</sup> It is also believed that in the presence of gender differences, allowing for negotiation may contribute to the gender wage gap. In the same survey of United States adults, women indicate that they are less likely than men to negotiate for their starting salary and are more likely than men to prefer environments in which negotiation is not allowed.<sup>2</sup> The belief that salary negotiation contribute to gender pay gaps in labor markets has lead some corporations to consider banning salary negotiation. One of the most notable examples of a Negotiation-Ban policy was the one enacted by former interim CEO of Reddit, Ellen Pao, in 2015.<sup>3</sup> The Negotiation-Ban has been lauded by some as being an effective policy to reduce gender wage disparities (Kray, 2015).

Proponents of Negotiation-Ban policies have pointed to a growing body of research suggesting that in many circumstances, women negotiate less often than men (Babcock and Laschever, 2003), may experience weaker gains from negotiation (Dittrich et al., 2014), and may even experience backlash from negotiation (Bowles et al., 2007). Negotiation-Bans are not without critics; some of the primary arguments against banning negotiations is that doing so may frustrate more skilled workers causing them to reduce their output and that if managers have implicit gender biases, Negotiation-Bans may not reduce and could even increase gender wage disparities.<sup>4,5</sup>

---

<sup>1</sup>Results from an MTurk Survey (n=317) that asked “In which company do you think the difference in worker earnings is the largest” (A company that allows/does not allow for salary negotiation). Sample is restricted to those that correctly answer attention checks.

<sup>2</sup>When asked whether they negotiated the starting salary at their most recent job, 37% of men (n=182) and 27% of women (n=133) said they did (two-sided t-test, p=.03). We exclude the two participants that indicated they have never been employed. When asked whether they would prefer an environment that allows for salary negotiation, 93% of men (n=183) and 87% of women (n=134) indicated that they would (two-sided t-test, p=.05).

<sup>3</sup>Other examples of companies or company affiliates that have publicly stated they do not allow salary negotiations include former Principle Recruiter for Google Engineering, Bob See, on Quora (<http://qr.ae/TUpHPj>) and The Guardian article by Communications Director, Carys Afoko, of SumOfUs (<https://www.theguardian.com/commentisfree/2017/aug/02/ban-individual-salary-negotiations-gender-pay-equality>).

<sup>4</sup>See, for instance, this blog post by consulting firm The Azara Group (<http://www.theazaragroup.com/banning-salary-negotiations-ellen-pao/>), this article in HR Magazine (<https://www.shrm.org/hr-today/news/hr-magazine/pages/0915-salary-Negotiation-Bans.aspx>), and Wharton Professor Bidwell quoted in this article by Mashable (<https://mashable.com/2015/04/06/ellen-pao-reddit-salary/>).

<sup>5</sup>One commonly seen suggestion by opponents of Negotiation-Bans is that, rather than removing the option to negotiate, women should be encouraged to enter into negotiations more often. Evidence from Exley et al. 2018, however, demonstrates that such advice is misguided; pushing women into negotiation does not improve their outcomes. Women that refrain from negotiation are harmed by being forced into negotiation.

Despite the significant attention Negotiation-Bans have received by media, business professionals, and scholars alike, to our best knowledge no study has formally tested the implications of a Negotiation-Ban. The finding that women frequently benefit less than men in environments with negotiation does not necessarily imply that a Negotiation-Ban will reduce gender inequality in compensation. If managers hold implicit gender biases, as critics of Negotiation-Ban policies suggest, Negotiation-Bans could have no or negative effects on creating more equitable earnings distributions within a workplace. To determine whether Negotiation-Ban policies may reduce gender inequality in compensation, this study provides the first examination of outcomes from a setting where individuals are allowed to negotiate to the counterfactual outcomes from a setting where individuals are banned from negotiating.

We conduct a laboratory experiment in a two-worker environment with repeated interaction in which workers within a workgroup perform separate tasks that differ in productivity. Participants assigned to the role of worker complete their assigned real-effort task to secure earnings for their group. Participants assigned to the role of manager view their worker characteristics and output and must decide how to split earnings between the two workers in their group. Manager earnings are purely a function of their workers' output and thus are not affected by their compensation decisions; however, the repeated nature of the experiment means that manager compensation decisions may affect their future earnings via changes in worker output. We vary whether workers can engage in private negotiation with their manager. In our Negotiation treatment, workers can choose to enter negotiations with their manager prior to the manager deciding worker compensations. By contrast, workers in our Negotiation-Ban treatment do not have the option of entering negotiations.

We design our experiment to rigorously test the potential for a Negotiation-Ban to be an effective policy at reducing gender wage gaps. As such, we designed our experiment to increase the likelihood of total worker output decreasing under a Negotiation-Ban and to increase the likelihood of managers' implicit biases affecting their compensation decisions under a Negotiation-Ban. The heterogeneity in task productivity helps to accomplish both goals. The heterogeneity creates more room for manager's implicit biases to affect their compensation decisions. Further, heterogeneity in task productivity creates a higher compensation demand among high-productivity task workers and may lead to diminished productivity when they are unable to negotiate for higher pay. Finally, the repeated matching in our environment ensures managers are incentivized to compensate workers in a manner that will maximize total output.

We show that, contrary to what has been suggested by critics of Negotiation-Bans, worker output does not decrease for either worker in the presence of a Negotiation-Ban. This suggests that critics' concerns about Negotiation-Bans leading to decreased effort among high-productivity workers are likely unfounded. As further evidence, we also show that dispersion in the share of output produced by high-productivity task workers within worker groups is the same between treatments.

In examining manager compensation decisions, we show that managers indeed experience a trade-off between their own fairness ideals and the desire to maximize total worker output. As such, the majority of compensation

decisions are shaded in favor of the high-productivity worker but do partially account for the difference in relative task difficulty. Unlike previous studies that have examined multi-worker environments (for instance, see [Gross et al. \(2015\)](#) and [Bolton and Werner \(2016\)](#)), this paper is the first to ask how wage compression differs in environments with versus without negotiation. We find inequality of both the dollars earned by workers (pooled over both tasks) and the share of earnings received by high-productivity task workers are both lower when negotiation is banned. Combining this finding with the result that dispersion in output share is the same across treatments, our study suggests that banning negotiation leads to increased wage compression.

We further demonstrate that the reduction in earnings inequality is largely coming from groups in which the high-productivity worker is male. Focusing on individual returns to negotiation, we show that this result can be explained by the fact that only male workers assigned to the high-productivity task benefit from entering negotiation; female workers on the high-productivity task and all low-productivity task workers do not experience any return from entering negotiation. Using coded chat data, we determine that this gap is partially due to the fact that only high-productivity workers benefit from threatening to reduce output and that male workers are much more likely to use threats. However, even after controlling for a broad assortment of negotiation styles, we still find gender differences in returns to negotiating for high productivity workers.

Our findings demonstrate that, in contrast to critics' concerns about Negotiation-Ban policies, under such a ban gender gaps in earnings decrease and productivity does not decrease. As such, this study provides encouraging evidence to support the possibility of Negotiation-Ban policies being effective at reducing compensation inequalities that emerge from gender differences in returns to negotiation. Furthermore, we demonstrate that eliminating the ability of high productivity workers to leverage their greater bargaining power, such policies have the added benefit of reducing earnings inequalities in environments where workers perform equally demanding tasks that differ in productivity.<sup>6</sup>

The rest of this chapter proceeds as follows. Subsection 1.2 provides a detailed description of our experimental design. We describe the data in Subsection 1.3, detail the results in Subsection 1.4, and discuss factors that may affect differences in returns to negotiation in Subsection 1.5. Finally, in Subsection 1.6 we conclude.

## 1.2 EXPERIMENTAL DESIGN

We design our experiment using ZTree ([Fischbacher, 2007a](#)). In order to create room for managers to have subjective assessments of worker performance, workers in our experiment are randomly assigned to perform one either a high-productivity task and a low-productivity task. To ensure managers' compensation choices affect their future earnings, our experimental design uses repeated matching between managers and workers. The remaining paragraphs in this section describe our experimental design in detail.

---

<sup>6</sup>One example of such an environment is a workplace where workers sell similar or identical products to different demographics. For instance, pharmaceutical representatives with sales areas that differ in the how urban versus rural they are.

Participants in our experiment are evenly split between Managers, Green Workers (low-productivity task), and Blue Workers (high-productivity task) at the beginning of the experiment. We use colors to refer to the two types of workers to avoid experimenter demand effects. Each participant is placed into a group consisting of one Manager, one Green Worker, and one Blue Worker. Participants stay in their role and group for the duration of the experiment.

In our experiment, we use a between-subject design consisting of a “Negotiation” treatment and a “Negotiation-Ban” treatment. In the Negotiation treatment, workers have the opportunity to initiate a negotiation with their manager whereas in the Negotiation-Ban treatment, there is no opportunity to negotiate. In both treatments, Participants face five work rounds. In the Negotiation treatment, these rounds consist of a production stage, a negotiation stage, and a payment-selection stage. In the Negotiation-Ban treatment, these rounds consist only of a production stage and a payment-selection stage.

For each round, workers have 120 seconds to perform their work task during the production stage. Each unit of output produced by a worker secures \$0.025 for the manager and \$0.05 for joint worker earnings. During the payment-selection stage, the manager views both worker output and worker characteristics (as provided at the beginning of the experiment) and must decide how to split joint worker earnings between the two workers in their group. In the negotiation treatment, the payment-selection stage is preceded by a negotiation stage in which workers can pay \$0.05 out of their personal earnings to initiate a negotiation where they may engage in free form text chat privately with their manager for up to three minutes.

The instructions provided to participants contain information on each role, the two different work tasks, and – in the negotiation treatment – a description of how negotiations take place. Instructions also contain screenshots of each work task and the manager decision screen. At the beginning of the experiment, prior to having read the instructions or learned their assigned role, participants complete a brief survey asking for their year in college, age, gender, major, and whether they attended high school in Pennsylvania. These answers are provided later to the manager in the workers’ group. Instructions are read out loud to guarantee that the structure of the experiment is common information. After reading instructions, participants learn their assigned role and workers are given 30 seconds to practice their assigned work task. In the subsequent paragraphs, we discuss each aspect of the experiment in more detail.

### **1.2.1 Production Stage and Work Tasks**

At the beginning of the experiment, participants are assigned to either the Green Worker role or the Blue Worker role. The two worker roles differ in the task they must perform during each production stage. Green workers perform a Slider task ([Gill and Prowse, 2012](#)) while Blue workers perform a Typing task. We ex-ante expected the Typing task to be a higher productivity task than the Slider task. We chose to have workers perform two different tasks to create room for managers to form subjective evaluations about relative worker performance. This wiggle room increases the possibility for us to observe differences in worker earnings resulting from negotiation, and for it to potentially vary with worker characteristics such as gender. Each unit of output produced by a worker increases manager earnings in

their group by \$0.025 and joint worker earnings by \$0.05. We selected these earnings amounts because they allow managers the option to have all group members earn the same if they select a 50/50 split between the two workers. Additionally, having tasks that differ in productivity but not returns to a unit of output also allows for subjectivity in terms of what split of earnings is fair.

During each production stage, workers have 120 seconds to complete their assigned task. Workers assigned to the Slider task see a screen displaying 48 sliders offset from one another. To complete one unit of output, the workers must position the marker of the slider at exactly 50. To position a marker, workers must click on the marker with their mouse and drag it horizontally to the desired location. Information on the present location of the marker is only updated after releasing the mouse. Total output for the Slider-worker in a given period equals the number of sliders for which the marker is correctly positioned at 50 by the end of the production stage.

Workers assigned to the Typing task (Typing-workers) are shown a screen with two boxes located side by side. Initially, the box on the right contains a letter highlighted in red and the box on the left is blank. After a Typing-worker correctly types the letter in the box on the right, the letter turns black and a new letter highlighted in red appears in the left box. Correctly typing the letter in the left box completes a letter pair. After a letter pair is completed the screen is reset; the worker is given a new letter highlighted in red in the right box and the left box is blank. Total output for the Typing-worker in a given period equals the number of letter pairs completed by the end of the production stage.

Managers view updates of their workers' progress during the production stage. Specifically, managers are shown a progress screen updated every 10 seconds. The progress screen provides the survey answers for the two workers in their group, current output by each worker in the given period, current total output in the given period, current manager earnings, and current joint worker earnings.

### **1.2.2 Negotiation Stage**

In the Negotiation treatment, after completing the production stage, workers are given the opportunity to pay \$0.05 to privately negotiate with the manager in their group. We chose to make negotiation costly to ensure workers do not simply initiate negotiations to pass time and also to better mirror the fact that workplace negotiations are costly in terms of time-required, risk, and emotional energy.

If a worker initiates a negotiation, they will have up to three minutes to engage in free-form chat privately with the manager in their group. Free form chat allows for a multitude of negotiation tactics. Once a worker is done negotiating, they have an option to leave the chat window and are taken to a waiting screen until after both the negotiation stage and the payment-selection stage is done. Managers do not have the option to exit the chat window. Workers that do not enable chat are immediately taken to this waiting screen. Doing this prevents workers from learning about whether or for how long the other worker in their group is chatting.

If both workers select into negotiation, the manager is given a separate chat window for each worker. If a worker decides not to negotiate, the manager is not given a window for that worker. Once both workers have finished negotiating or if neither worker initiates a negotiation, the manager is immediately taken to the payment-selection stage.

### **1.2.3 Payment-Selection Stage**

During the payment-selection stage, Managers see their earnings for the period, joint worker earnings, individual characteristics (from survey responses) of their two workers, the workers' assigned roles, and the output of each worker in their group. Using a payment selection bar, the Manager selects a value  $P$  between 0 and 1 in increments of 0.01 to determine what share  $P$  of joint worker earnings will be paid to the Slider-worker. The remaining  $(1-P)$  share is given to the Typing-worker. Managers in the negotiation treatment also have an option of reviewing their negotiations for the given period. Once all managers complete the payment selection stage, all participants are shown a summary screen of each workers' output in the given period and the earnings of each member in their group.

### **1.2.4 Follow-up Activities**

After participants complete the five rounds of the main experiment, they complete a modified dictator game task that allows us to identify any explicit taste-based discrimination. Participants are matched into new groups of three for this task. Participants are no longer assigned roles in this task. Each participant sees the survey responses for the two other individuals in their group and must decide how to allocate \$5 between their two group members. For each group, we randomly select one group member's decision to implement. The selected group member earns \$2.50 and the other two group members earn the specified split.

After completing the modified dictator task, participants complete a survey that asks about their perceptions of task difficulty. This measure allows us to ask whether Managers' beliefs affect their compensation decisions. Following this, we ask participants what they believe a fair split would be for three different output levels: 10 sliders and 90 letter-pairs; 25 sliders and 75 letter-pairs; 40 sliders and 60 letter-pairs. Finally, participants complete a brief demographic survey.

## **1.3 DATA**

The study was advertised as a 90-minute decision-making study with a guaranteed minimum payment of \$6 and expected average earnings that would exceed this amount. No additional details were given about the study.<sup>7</sup> 252

---

<sup>7</sup>To facilitate achieving gender-balanced sessions, we created two separate studies for each session: one for male participants and one for female participants. Potential participants can only view studies they are eligible for. Thus, we were able to discreetly recruit equal numbers of men and women for each session.

undergraduate students participated in fourteen sessions at the Pittsburgh Experimental Economics Laboratory (PEEL). Each session included eighteen participants. We used PEEL's SONA system database to recruit student participants. From the main experiment, additional payments from follow-up activities, and the \$6 show-up fee, average earnings were \$22.70.

For each session males and females were (unbeknownst to them) signing up on two separate study lists, this helped us to secure a relatively gender balanced sample. The average percentage of female participants across sessions was 48%. The average age of participants was 19.69, with 99% ranging from age 18 to 23. Most participants (46%) were majoring in Business or Social Sciences, followed by Engineering or Natural Sciences (35%), and other or undeclared majors (19%). The study consisted of 84 manager-worker groups. We observe 41 groups in the Negotiation treatment and 41 groups in the Negotiation-Ban treatment. For each group, we observe 5 rounds of data.

## **1.4 RESULTS**

In this section, we ask four main questions. First, pooling the data for our two treatments, we confirm that our tasks differ in productivity and ask how managers choose to distribute earnings between the high- and low-productivity-task workers. We then ask whether removing the option to negotiate affects the distribution of compensation. After observing differences in the distribution of compensation between treatments, we ask whether the individual worker benefits from engaging in negotiation, and whether this varies with the worker's task. Finally, we ask whether manager compensation decisions and the effect of removing negotiation differ by worker gender.

When considering compensations, managers face two key considerations. First, managers may wish to pick the compensation scheme that they deem to be fair given the output of the high- and low-productivity task worker in their group. Second, managers may consider how their compensation choice will affect worker output in the subsequent periods. Given that workers assigned to the high-productivity task can secure greater manager earnings, this second motivation may cause managers to select higher compensation for high-productivity-task workers by more than they otherwise would. This could in turn lead to managers making compensation choices differently in the last period (when the need to incentivize future performance is removed) compared to the first four periods. Thus, we anticipate decision behavior by managers may exhibit significant end-game effects. Given that we are primarily interested in work environments in which workers will have continued interaction with their employer, we restrict much of the remaining analysis to only include the first four rounds of decision data.

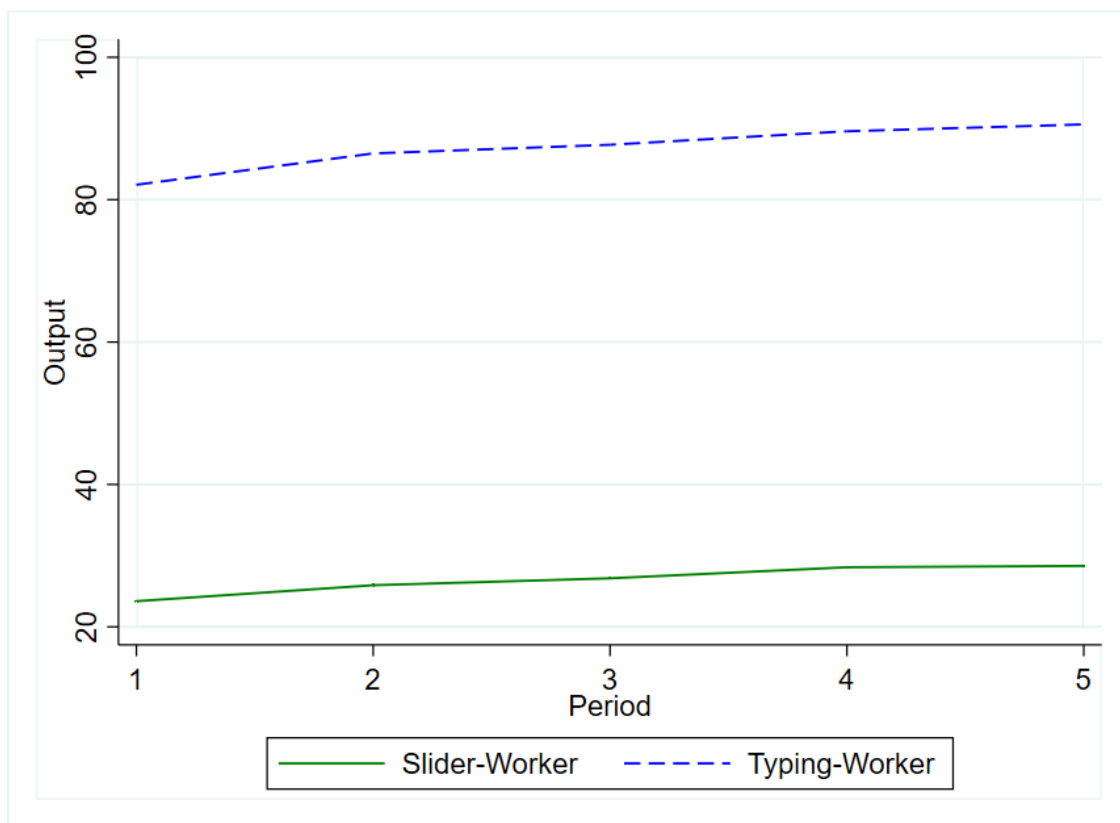
### **1.4.1 How do Managers Compensate Workers?**

In this subsection, we examine worker output and the corresponding manager compensation choices pooling data from the Negotiation and Negotiation-Ban treatments. We begin by comparing average output between the two tasks. As



shown in Table 1.1, we confirm that with an average output of 87.3 letter-pairs per period versus an average output of 26.6 sliders per period, the typing-task is indeed a higher productivity task than the slider-task. Given that the typing-task is likely to be more familiar to participants than the slider task, one concern may be that this productivity difference falls over time; however, as demonstrated in Figure 1.1, the gap in output between the two tasks remains relatively constant. Finally, in response to critics' concerns that a Negotiation-Ban may reduce output among high productivity workers, we compare worker output between treatments. Figure 1.2 shows clearly that this is not the case. If anything, absent the ability to negotiate, worker output is slightly higher.

Figure 1.1: Output over Time by Worker Task



If managers care solely about output, we would expect to see average Typing-workers earning 3.28 times that of the average Slider-worker. Instead, Table 1.1 shows that the mean period earnings for a Typing-worker is \$3.35 while the mean period earnings for a Slider-worker is \$2.45. That is, Typing-workers on average earn 1.37 times that of Slider-workers. While this is substantially less than the difference we observe in output, it is transparent that the Typing-worker is compensated for greater productivity. Using a two-sided t-test, we reject that Slider-workers and Typing-workers have the same earnings ( $p=0.00$ ).

Figure 1.2: Output over Time by Worker Task and Treatment

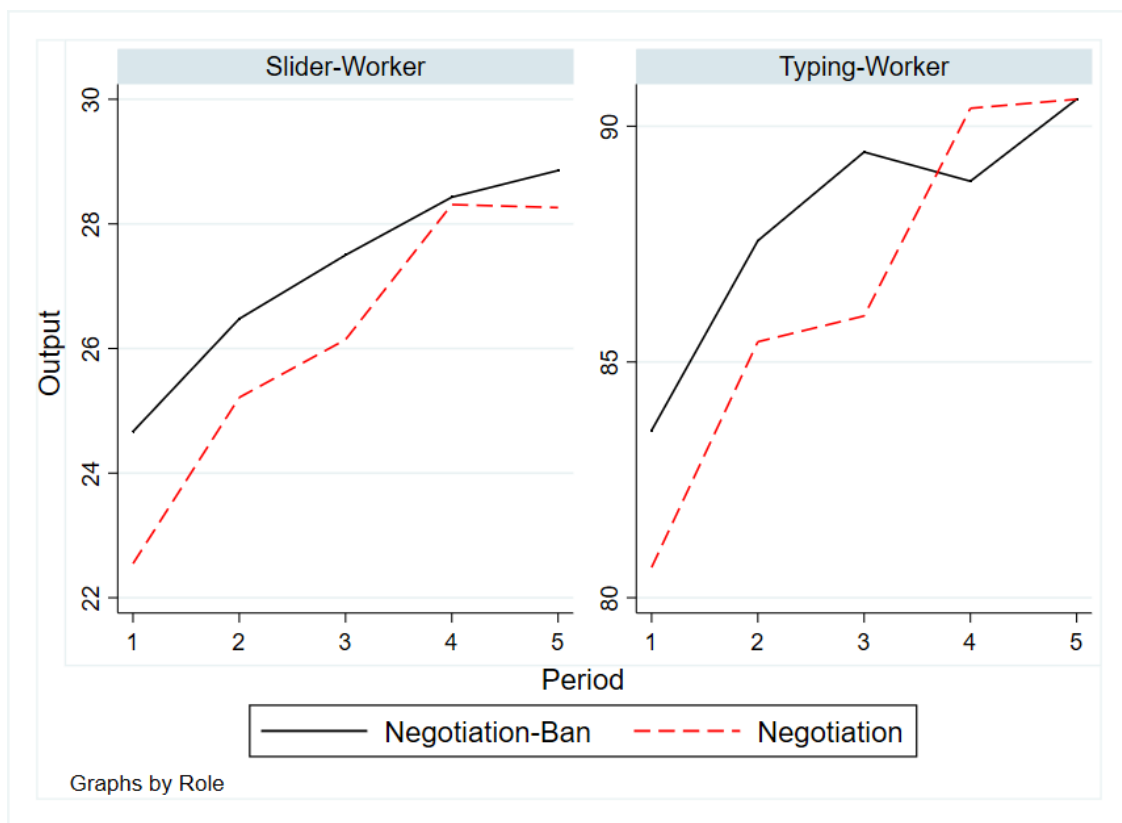


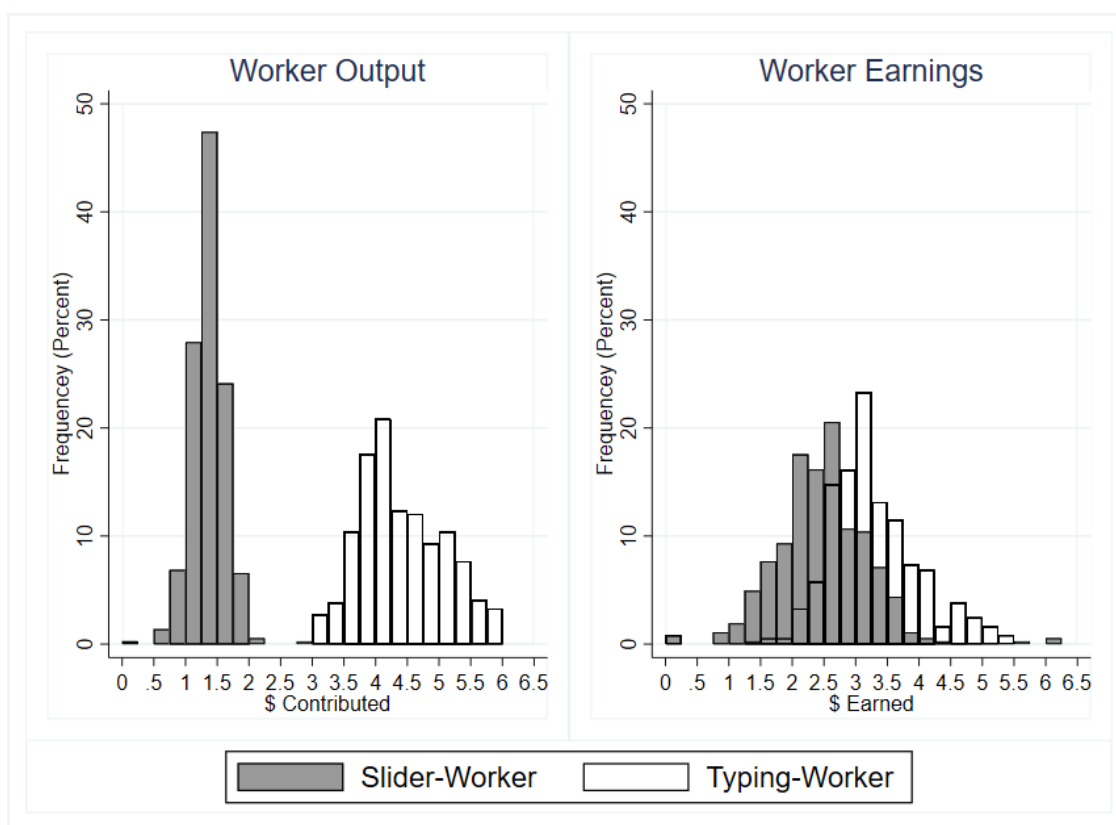
Table 1.1: Summary Statistics

	Pooled	Negotiation	Negotiation-Ban
Letter-Pairs	87.3 (13.8)	86.6 (14.1)	88.0 (13.6)
Sliders	26.6 (5.1)	26.1 (5.4)	27.2 (4.7)
Typing-Worker Earnings	\$3.25 (0.76)	\$3.25 (0.79)	\$3.25 (0.72)
Slider-Worker Earnings	\$2.45 (0.71)	\$2.39 (0.74)	\$2.51 (0.68)
Joint Worker Earnings	\$5.70 (0.78)	\$5.64 (0.80)	\$5.76 (0.75)
N	420	210	210

Notes: Standard deviations are in parentheses. Joint Worker Earnings is the sum of Slider and Typing-worker earnings within a group.

To further illustrate the differences between what workers produce and what they receive, Figure 1.3 compares the distributions of each workers' contribution to joint worker earnings in a given period (left panel) to the distributions of each workers' compensation in a given period (right panel). If workers are paid just what they contribute, the left and right panels should be identical. In contrast, under full wage compression, the earnings distributions for the Slider and Typing-worker should be identical. Figure 1.3 shows that earnings are somewhere in between the two extremes. While the worker contribution distributions are disjoint except for a very small number of zero-output observations, there is partial overlap in the distributions of the Typing- and Slider-worker earnings. In other words, we find evidence of partial wage compression between the two workers.

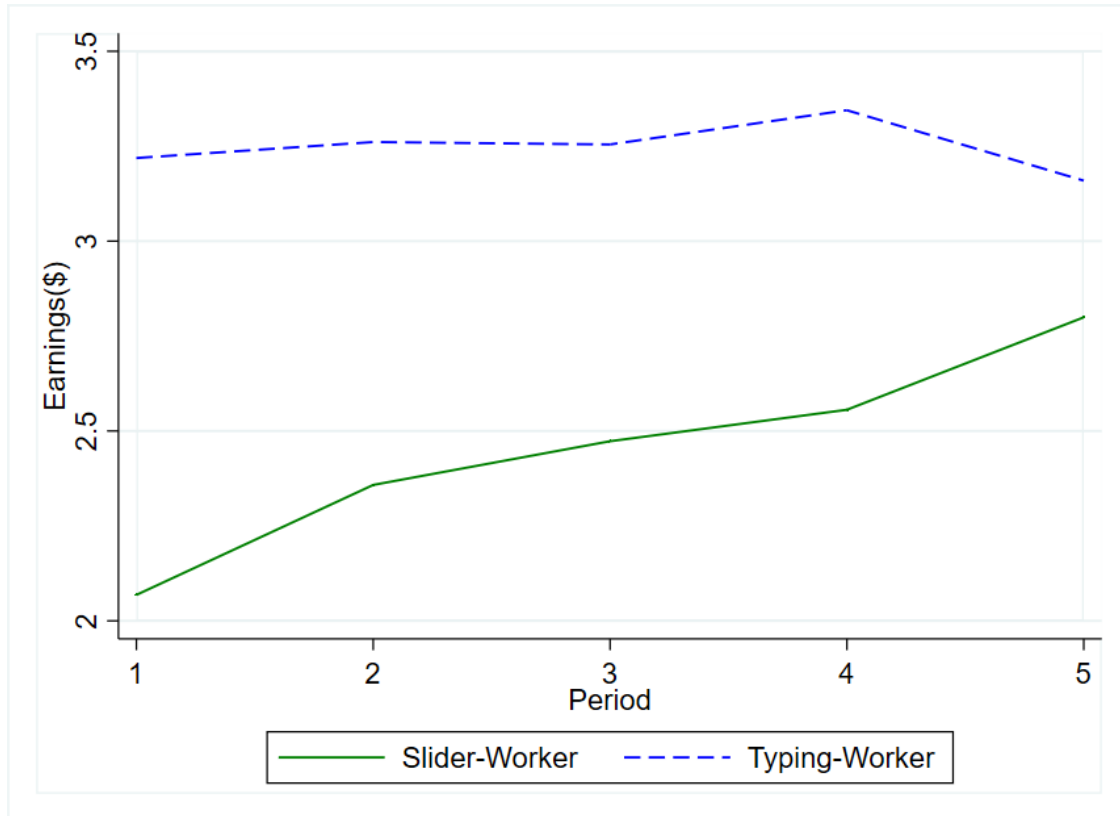
Figure 1.3: Distribution of Worker Output and Worker Earnings



Compensation decisions may vary over time, however, if it takes time for managers to learn about relative task difficulty or if end-game effects exist. Figure 1.4 shows a modest decrease in earnings differences over the first four periods with the largest increase occurring after the first period; however, the difference in compensation shrinks substantially more in the final period even though there is no similar decline in Typing-worker output in the final period (Figure 1.1). We see a similar pattern when we instead look at the share of joint worker earnings being produced by the Typing-worker and compare it to the share of joint worker earnings being allocated to the Typing-worker in Figure

1.5. These results suggest that while there are slight learning effects by managers, end game effects are much more pronounced. Given the substantial end-game effects, we restrict our remaining analysis in this study to the first four periods except where noted otherwise.

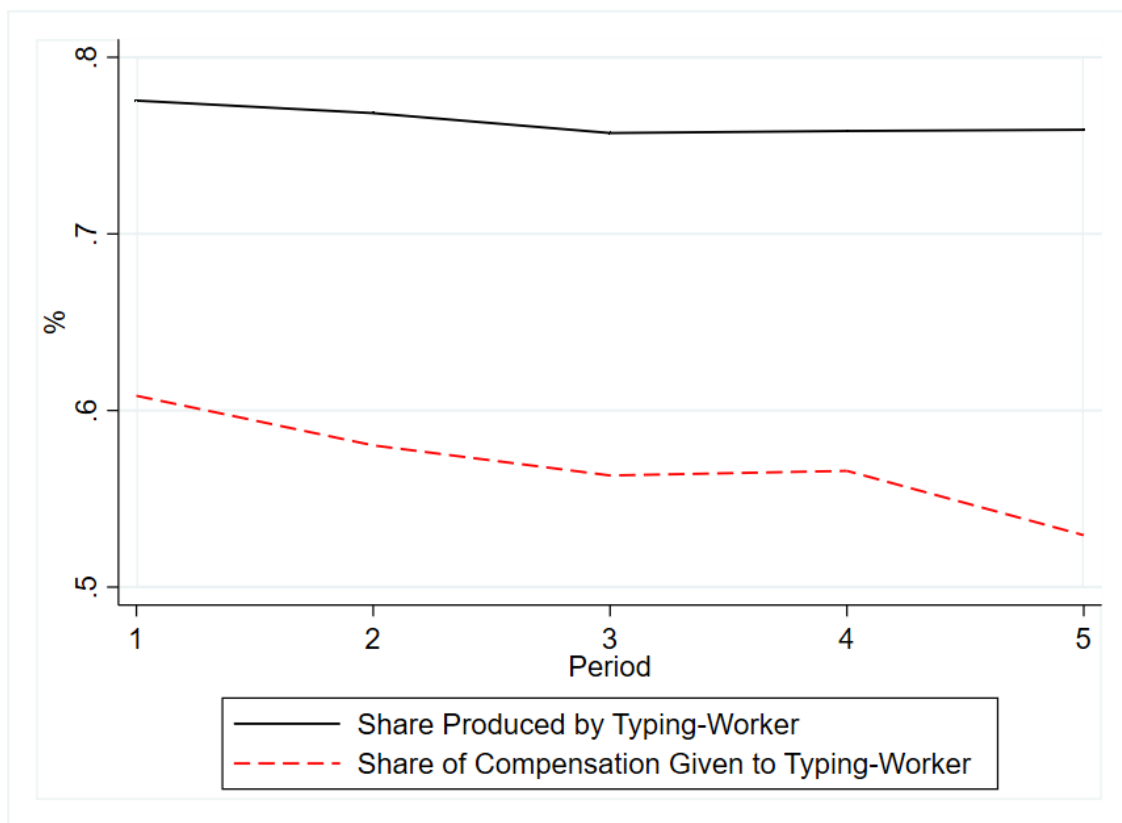
Figure 1.4: Earnings over Time by Worker Task



One of the primary differences between the first four periods and the last period is that managers no longer face the possibility that their compensation choice will affect future productivity of their workers. As such, the compensation decision in the final period is likely to be more reflective of what the manager perceives to be a “fair” allocation of joint worker earnings between the two workers. The fact that we see a decline in Typing-worker earnings and earnings share in the final period suggests that incentive concerns lead managers to compensate Typing-workers more than their fair share because Typing-workers have higher potential productivity. As further evidence of this, we use managers survey responses to calculate their perceived difficulty adjusted output share produced by the Typing-worker. We show in the appendix that the majority of compensation decisions are greater than this difficulty adjusted share and that the difference is considerably smaller in the fifth period.

To summarize, we confirm that the Typing-task is a higher productivity task than the Slider-task and find that in an environment where managers decide how to allocate earnings between two workers that differ in their productivity,

Figure 1.5: Relative Production and Compensation over Time



managers engage in some wage compression (as shown in Figure 1.3). This wage compression, however, is less than what managers would select if they were to compensate workers solely based on perceived effort (as shown in the appendix). Given that wage compression increases in the last period, this is likely due to incentive concerns faced by the manager. Furthermore, consistent with [Gross et al. \(2015\)](#) and [Bolton and Werner \(2016\)](#), the fact that Slider-worker output increases over time suggests Slider-workers accept the unequal distribution of earnings at least to the extent that they do not reduce effort due to discouragement or dissatisfaction with being a lower earner.<sup>8</sup>

#### 1.4.2 What are the Institutional Effects of a Negotiation-Ban?

Having confirmed that worker productivity differs by task and that our environment produces partial wage compression, we ask whether the imposition of a Negotiation-Ban affects the relative distribution of compensation by worker task. While Typing-workers that enter into negotiation increase their earnings, this does not necessarily imply that imposing a ban on negotiation will affect the relative earnings of Typing versus Slider-workers. As an example, suppose that when workers have the option to negotiate, managers provide a boost to compensation above what they believe is fair only to the Typing-workers that demand higher wages via negotiation; however, in an environment without negotiation, managers boost compensation only for those Typing-workers that they believe would demand higher earnings if negotiation were available. If manager beliefs about who would negotiate are sufficiently accurate, then we would not expect to see the enactment of a Negotiation-Ban affecting the relative earnings distribution.

To answer whether a Negotiation-Ban affects the relative earnings distribution, we calculate the Gini coefficient for each treatment. We start by pooling per-period earnings for both the Slider-workers and the Typing-workers. Table 1.2 shows that the Gini coefficient for pooled earnings is 0.17 in the Negotiation treatment and 0.15 in the Negotiation-Ban treatment— an 11.7 percent decrease that is significant at the 5% level. This reduction in inequality is equivalent to imposing an 11.7 percent proportional tax on workers in the Negotiation treatment and redistributing the tax revenue as equal sized amounts to each individual worker ([Aaberge, 1997](#)). While we observe a reduction in earnings inequality in the Negotiation-Ban treatment, it is possible that this inequality reduction could be the result of reduced inequality in output. Thus, we next turn to looking at the inequality in the share produced by the Typing-worker and the share earned by the Typing-worker. The benefit of these latter measures is that they control for who the other worker is paired with. Table 1.2 shows that we observe no significant decrease in the share produced. We do, however, observe that the Gini on share earned falls by a statistically significant amount from 0.11 to 0.088 moving from the Negotiation to Negotiation-Ban treatment. From this, we conclude that imposing a Negotiation-Ban leads to a reduction in compensation inequality even after accounting for output differences.<sup>9</sup>

---

<sup>8</sup>Mean Slider-worker output is consistent with that found in [Araujo et al. \(2016\)](#); however, improvement over time actually occurs at a steeper rate in our study. This suggests that, if anything, Slider-workers in our environment exert more effort than in a piece rate payment environment.

<sup>9</sup>An alternative method used for measuring inequality is to compare the earnings of the 90th percentile with those of the 10th percentile. Columns 3 and 4 of Table 1.2 show that this measure produces similar, albeit weaker, results.

Table 1.2: Gini Coefficient and 90-10 Ratio, By Treatment

	Gini: Negotiation	Gini: Ban	90-10: Negotiation	90-10: Ban
Pooled Earnings	0.17 (-0.008)	0.15** (-0.006)	2.2 (-0.11)	1.96* (-0.08)
Pooled Output	0.32 (-0.009)	0.31 (-0.007)	4.62 (-0.176)	4.39 (-0.131)
Share Produced by Typng Worker	0.035 (-0.006)	0.028 (-0.0016)	1.14 (-0.014)	1.13 (-0.009)
Share Earned by Typing Worker	0.11 (-0.0081)	0.088** (-0.0038)	1.54 (-0.078)	1.49 (-0.056)

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Stars indicate significance for two-sided t-test between treatments. Bootstrapped standard errors in parentheses. Observations restricted to first four periods. Pooled output includes both Slider task and Typing task output. Pooled earnings includes both Slider-worker and Typing-worker earnings.

To summarize, does a Negotiation-Ban matter for the relative distribution of earnings between workers? The answer is a clear yes. The imposition of a Negotiation-Ban leads to a substantial decline in compensation inequality among workers. Furthermore, given that we do not observe distributional changes in relative production, this reduction in inequality is not explained by any changes in output across treatments. Put differently, these results demonstrate that while wage compression exists in our environment, allowing workers to negotiate increases earnings inequality and thus reduces wage compression.

### 1.4.3 Why does a Negotiation-Ban Reduce Compensation Inequality?

Given our finding of reduced compensation inequality under a Negotiation-Ban, it is likely that managers are responding to workers attempts to negotiate when it is allowed. There are three primary ways in which engaging in negotiation may affect compensation inequality. First, Slider-workers may attempt to negotiate and experience backlash. Second, Typing-workers may experience positive returns from negotiating. Third, it is possible that both workers benefit equally conditional on entering negotiation, but Typing-workers enter more frequently. In this subsection, we begin by confirming that workers engage in negotiation. We then ask what the returns are to negotiation by worker task.

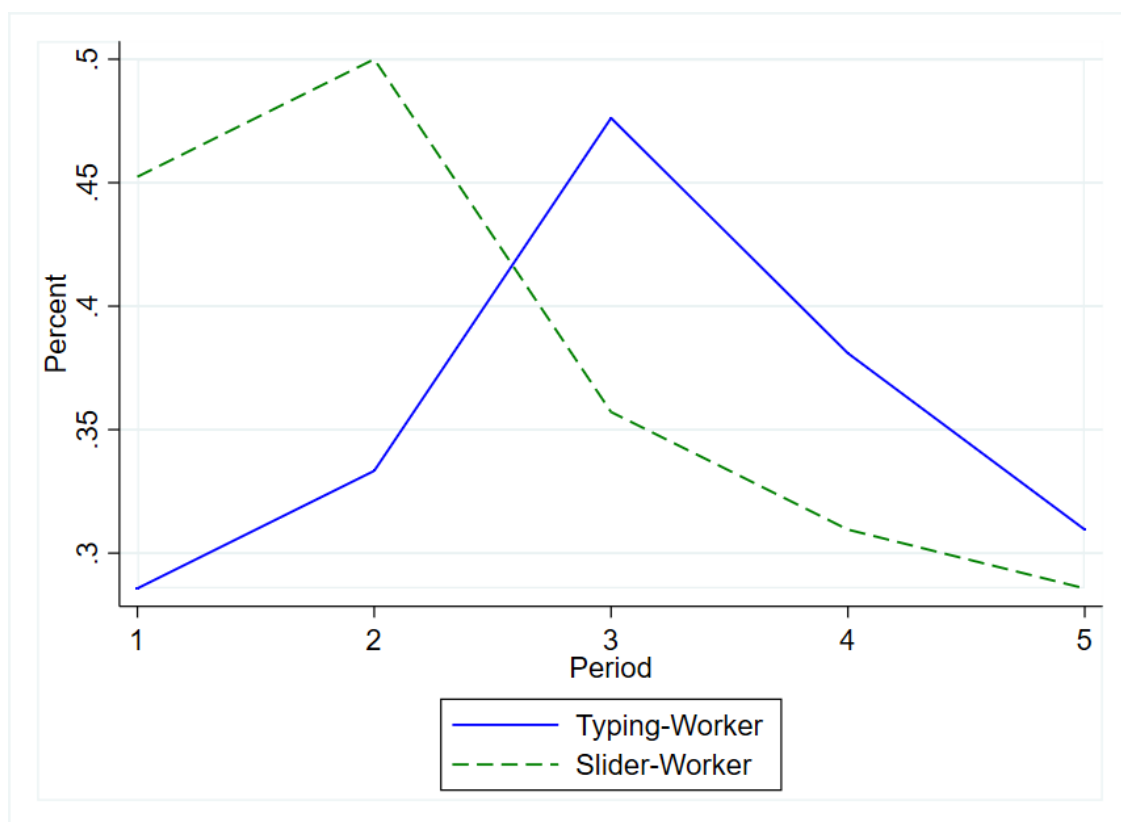
Looking at the data from the negotiation treatments (and restricting our attention to the first four periods), we find that workers enter negotiations 36% of the time. Although Typing-workers may have greater potential bargaining power, we do not observe substantial differences in the overall negotiation rate between the two types of workers.



Slider-workers enter negotiation 40% of the time and Typing-workers enter negotiation 37% of the time. Using a two-sided t-test, we fail to reject that these rates are different ( $p=0.50$ ).<sup>10</sup>

Figure 1.6 shows that the negotiation rates over time decrease in the 4th and 5th period. One explanation may be that managers and workers reach initial agreements that cover outcomes for subsequent periods (thus decreasing the need for subsequent negotiations). An alternative explanation is that workers anticipate returns from negotiating are decreasing over time. We do observe some differences in the timing of negotiations between worker types. Figure 1.6 also shows that negotiation rates for Slider-workers are initially higher than for Typing-workers and reach their peak earlier (in the second versus the third period). This difference in timing suggests Slider-workers and Typing-workers may utilize negotiation in different ways.<sup>11</sup>

Figure 1.6: Share of Workers who Initiate Negotiation over Time by Worker Task



Given that we observe both worker types opting in to negotiation, we next ask whether it pays off to negotiate and whether this differs by worker type. To do this, we run random effects regressions of workers' period earnings on the number of sliders completed in that period, the number of letter-pairs completed in that period, and an indicator

<sup>10</sup>We show in section 1.5, that while Slider- and Typing-workers enter negotiation at similar rates, they use very different negotiation tactics that reflect their different bargaining positions.

<sup>11</sup>This is confirmed in section 1.5.

variable that takes on a value of 1 if the worker opted to negotiate in that period. We include period fixed effects and cluster standard errors at the group level. We conduct this analysis separately by worker task.

As seen in column 1 of Table 1.3, Typing-workers that enter negotiation receive greater earnings in the periods that they enter. Typing-workers who enter negotiation increase their earnings by 23.4 cents. For negotiations to yield a positive return, however, it must be the case that the net return to negotiation (i.e., the return minus the 5 cent negotiation fee) is positive and significant. Testing for this, we find the net return, 18.4 cents is significantly different from zero ( $p=0.04$ ). In contrast to Typing-workers, Slider-workers do not benefit from initiating a negotiation. While we estimate that the coefficient on negotiation entry is 7.98 cents, we cannot reject that this is different from zero. Thus, only Typing-workers appear to benefit from initiating negotiations.

As further evidence that it is differential returns to negotiation entry driving the differences in inequality between institutions we regress Typing-worker earnings on output and negotiation entry pooling together both treatments. We add an indicator variable for the Negotiation-Ban treatment and control for negotiation entry by workers (note that the indicator variable for entry into negotiation will be 0 for all workers in the Negotiation-Ban treatment). Table 1.4 shows that for Typing-workers, the coefficient on Negotiation-Ban is small and insignificant whereas the coefficient on Entry is positive, significant, and of a similar magnitude to our estimates from Table 1.3. This suggests that Typing-workers that do not enter negotiation are paid similarly across treatments. As such, differences in earnings between treatments are primarily coming from positive returns to entry experienced by Typing-workers.

The result that only Typing-workers benefit from entering negotiations suggests that the reduction in compensation inequality under a Negotiation-Ban is likely due to Typing-workers no longer being able to leverage their high-productivity position to demand greater compensation from the manager. This demonstrates an additional benefit stemming from Negotiation-Ban policies; in workplaces where workers perform equally demanding tasks that differ in potential productivity, Negotiation-Bans can reduce earnings inequality between high and low productivity workers that result from high-productivity workers receiving preferential treatment. Such policy implications may be desirable for companies that wish to implement meritocratic fairness ideals.

#### 1.4.4 Gender

In addition to finding an overall difference in inequality following a Negotiation-Ban, it is also possible that the response to a Negotiation-Ban varies by gender. While it is generally found that women negotiate less frequently than men and secure smaller gains from negotiation on average, these results are highly context dependent ([Mazei et al., 2015](#)). In certain instances, the gender gap in returns to negotiation may even be reversed.<sup>12</sup> Given that in our environment the heterogeneity in tasks increases situational ambiguity about what is a “fair” split and given that workers are negotiating on behalf of themselves ([Bowles et al., 2005](#)), we hypothesize that males will experience a larger return to negotiation than females in our environment.

---

<sup>12</sup>For instance, when negotiating on behalf of others, women may fare better than men ([Bowles et al., 2005](#)).

Table 1.3: Per-Period Worker Earnings: Negotiation  
Treatment

	(1)	(2)
	Typing-Worker	Slider-Worker
# Letter-Pairs	0.0374*** (0.00252)	0.0134*** (0.00291)
# Sliders	0.0185** (0.00725)	0.0361*** (0.00827)
Negotiation Entry	0.234*** (0.0639)	0.0798 (0.0841)
Constant	-0.347 (0.253)	0.0755 (0.301)
N	168	168

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Standard errors are clustered at the group level and are shown in parentheses. Observations are restricted to the first four periods. Results are from random effects regressions of money given to the Typing-worker within a period in the negotiation treatment. All regressions include period fixed effects.

Table 1.4: Worker Compensation Pooling over Both  
Treatments

	(1)	(2)
	Typing-Worker	Slider-Worker
# Letter-Pairs	0.0350*** (0.00261)	0.0155*** (0.00265)
# Sliders	0.0143** (0.00579)	0.0381*** (0.00618)
Entry	0.236*** (0.0658)	0.0859 (0.0778)
Negotiation-Ban	0.00524 (0.105)	0.113 (0.106)
Constant	-0.0273 (0.251)	-0.179 (0.264)
N	336	336

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Observations are restricted to first four periods of data. Standard errors are clustered at the group level and are shown in parentheses. Results are from random effects regression on per-period worker earnings. Entry is a binary variable that equals one if the worker enters negotiation in the given period. Negotiation-Ban is a binary variable that equals one in the Negotiation-Ban treatment. All regressions include period fixed effects.

Absent negotiation, differential treatment of workers by gender can still occur however as a result of implicit biases held by managers tasked with selecting worker salaries. Even if males fare better in the negotiation treatment, it is not clear that a policy that bans negotiation will necessarily lead to reduced gender differences in compensation. Thus, we also test the implications of a Negotiation-Ban on gender differences in earnings by comparing outcomes for mixed-gender groups in the Negotiation treatment to outcomes in the Negotiation-Ban treatment.

For all analysis in this section, we restrict our attention to only mixed-gender groups. The rest of this section proceeds as follows, in section 1.4.4.1 we begin by providing an overview of the data for our mixed-gender group subsample. Next, in subsection 1.4.2.2, we ask whether differences in the distribution of compensation between treatments vary by the gender composition of worker groups. After observing the treatment effect of banning negotiation on compensation distributions differs by the gender composition of worker groups, we ask in subsection 1.4.2.3 whether returns to negotiation differ by worker gender.

Our primary goal in this subsection is to test whether a Negotiation-Ban reduces the gender differences in overall earnings that results from negotiation. We find that a Negotiation-Ban reduces the earnings advantage of the Typing-worker and does so more for males than for females. As a result, imposing a Negotiation-Ban reduces compensation inequality more in male high-productivity worker groups. We conclude that a Negotiation-Ban may be an effective policy at reducing gender earnings gaps that result from negotiation.

**1.4.4.1 Overview of Mixed-Gender Group Subsample** Our data consist of 17 same-gender worker groups and 67 mixed-gender worker groups.<sup>13</sup> By treatment, we observe 36 mixed-gender worker groups (7 same-gender) for the Negotiation treatment and 31 mixed-gender worker groups (11 same-gender) for the Negotiation-Ban treatment. Focusing on the mixed-gender worker groups, we observe 19 Male Typing – Female Slider groups and 17 Female Typing – Male Slider groups in the Negotiation treatment. We observe 17 Male Typing – Female Slider groups and 14 Female Typing – Male Slider groups in the Negotiation-Ban treatment.

Recall that managers were able to observe four characteristics about the workers in their group in addition to worker gender: age, major, year in school, and whether the worker attended high school in Pennsylvania. Before turning to our primary data analysis, we first note that there are no gender differences in these other measures. The average age for male workers is 19.69 and for female workers is 19.83 (two-sided t-test,  $p=0.49$ ). The share of male workers that attended high school in Pennsylvania is 41% and the share of female workers is 45% (two-sided t-test,  $p=0.66$ ). The share of male workers that are freshman, sophomore, and junior or above is 32.9%, 23.5%, and 43.5% respectively. For female workers, the distribution is 20.5%, 31.3%, and 48.2%. Using a Fischer's exact test, we are unable to reject the hypothesis that these two distributions are the same ( $p=0.17$ ). For major choice, 36.5% of male workers are Natural Science or Engineering majors, 50.6% are Business or Social Science, and 19.3% are Other or

---

<sup>13</sup>To increase the likelihood of having male-female worker groups while being discreet and still allowing for quasi-random assignment of roles and groups, we used seating cards to assign participants to seats and drew from two separate shuffled stacks of cards – one for men and one for women.

Table 1.5: Summary Statistics: Mixed Gender Pairs

	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled	Pooled	Negotiation	Negotiation	Ban	Ban
	Male	Female	Male	Female	Male	Female
	Typing	Typing	Typing	Typing	Typing	Typing
# Letter-Pairs	83.7	87.5	83.71	84.3	83.8	91.4
	(14.2)	(12.4)	(15.2)	(12.7)	(13.0)	(11.0)
# Sliders	24.9	27.5	23.9	26.8	26.1	28.2
	(4.6)	(5.3)	(4.0)	(5.9)	(5.0)	(4.42)
Typing-Worker	3.12	3.32	3.25	3.20	2.97	3.46
Earnings	(0.71)	(0.69)	(0.80)	(0.75)	(0.57)	(0.60)
Slider-Worker	2.32	2.43	2.13	2.36	2.53	2.52
Earnings	(0.64)	(0.63)	(0.66)	(0.57)	(0.56)	(0.69)
Joint-Worker	5.43	5.74	5.38	5.56	5.49	5.98
Earnings	(0.77)	(0.70)	(0.85)	(0.70)	(0.67)	(0.62)
N	144	124	76	68	68	56

Notes: Observations are restricted to first four periods of data and mixed gender pairs only. Standard deviations are in parentheses. Joint Worker Earnings is the sum of Slider and Typing-worker earnings within a group.

Undeclared. For female workers, 37.4% are Natural Science or Engineering majors, 43.4% are Business or Social Science, and 19.3% are undeclared. Using a Fischer's exact test, we are unable to reject the hypothesis that these two distributions are the same ( $p=0.47$ ).

Table 1.5 shows that Female Typing-workers perform slightly better on average than Male Typing-workers. In contrast, Female Slider-workers perform slightly worse than Male Slider-workers.<sup>14</sup> As a result, average joint worker earnings are lower in the Male Typing, Female Slider groups resulting in slightly lower earnings for these workers. Finally, we note that men and women are not equally likely to initiate negotiations. The share of male Typing-workers who initiate negotiations is 43% compared to 31% for female Typing-workers (one-sided t-test,  $p=0.061$ ). Similarly, male Slider-workers initiate negotiations 49% of the time compared to 37% for female Slider-workers (two-sided t-test,  $p=0.079$ ). This is consistent with previous literature that finds women tend to enter negotiation at lower rates than men (Babcock and Laschever, 2003). We can for neither gender reject that negotiation entry rates are independent of worker task ( $p=0.54$  for males and  $p=0.45$  for females).

**1.4.4.2 Do the Institutional Effects of a Negotiation-Ban Differ by Gender?** To answer whether the effects from imposing a Negotiation-Ban on compensation inequality differ by the gender composition of worker groups, we calculate the Gini coefficient and 90-10 ratio on relative output and relative earnings for each treatment and gender pairing separately.<sup>15</sup> We start by looking at male Typing-worker groups. Table 1.6 shows that Gini coefficient on the share produced by the Typing-worker is not significantly different between treatments. Thus, dispersion in output among male Typing-workers is similar between treatment. Looking next at the share allocated to the Typing-worker, we observe that for male Typing-worker groups, the Gini coefficient falls from 0.12 in the Negotiation treatment to 0.082 in the Negotiation-Ban treatment and that this difference is significant at the 1-percent level. This suggests that inequality in relative earnings decreased in the Negotiation-Ban treatment. Taken together, the similar level of output inequality between treatments paired with the decrease in earnings inequality for male workers suggests that the earnings advantage for male Typing-workers is higher in the Negotiation treatment relative to the Negotiation-Ban treatment.<sup>16</sup>

We do not observe the same pattern for female Typing-worker groups. As shown in Table 1.6 the Gini coefficient on the share produced by the Typing-worker falls from 0.034 in the Negotiation treatment to 0.024 in the Ban treatment. This difference is significant at the 1-percent level. The Gini coefficient on the share earned is not significant. Taken together, the reduction in output inequality paired with no reduction in income inequality suggests that, if anything, the earnings advantage for female Typing-workers is lower in the Negotiation treatment relative to the Ban treatment.

<sup>14</sup>Intriguingly, this difference is largest in the Negotiation-Ban treatment. While male Typing-workers have similar output levels across both treatments, female Typing-workers' output is higher in the Negotiation-Ban treatment.

<sup>15</sup>The results for the 90-10 ratio produce similar results to we observe for the Gini coefficient. Thus, we focus our discussion only on the Gini. Both the Gini coefficient and the 90-10 ratio are reported in Table 1.6.

<sup>16</sup>When considering the 90-10 ratio, the results are different, but produce a similar conclusion. Output inequality increases by a significant amount, and there is no significant change in earnings inequality. Combined, these two effects also imply that the earnings advantage for male Typing-workers is higher in the Negotiation treatment relative to the Negotiation-Ban treatment.

Table 1.6: Gini Coefficient and 90-10 Ratio, By Treatment and Gender Composition

	Male Typing-Worker		Female Typing-Worker	
	(1)	(2)	(3)	(4)
	Negotiation	Ban	Negotiation	Ban
Panel A. Gini Coefficient				
Share Produced	0.035	0.031	0.034	0.024***
by Typing-Worker	(0.011)	(0.002)	(0.0033)	(0.0020)
Share Earned	0.12	0.082***	0.088	0.093
by Typing-Worker	(0.013)	(0.006)	(0.0086)	(0.0056)
Panel B. 90-10 Ratio				
Share Produced	1.11	1.15**	1.17	1.12*
by Typing-Worker	(0.010)	(0.014)	(0.029)	(0.014)
Share Earned	1.52	1.36	1.46	1.53
by Typing-Worker	(0.11)	(0.076)	(0.077)	(0.059)

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Stars indicate significance for two-sided t-test between treatments. Bootstrapped standard errors in parentheses. Observations are restricted to mixed-gender groups and the first four periods only. Pooled output includes both Slider task and Typing task output. Total output is the sum of Slider task and Typing task output. Pooled earnings includes both Slider-worker and Typing-worker earnings.



**1.4.4.3 Why do the Effects of a Negotiation-Ban Differ by Gender Composition?** Differences in compensation and output inequality by worker gender compensation, suggest that male typing-task workers experience a reduced earnings advantage following a Negotiation-Ban and female typing-task workers experience (if anything) an increased earnings advantage following a ban. This suggests that male and female workers may experience different returns to negotiation. In this subsection, we explore this possibility.

To formally test whether there exist gender differences in returns to negotiation, we use a random effects regression, clustering standard errors at the group level, to estimate the return to negotiation in a given period by worker task and gender. As before, we restrict our attention to the first four periods. We begin by confirming that we observe similar returns to negotiation in our subsample as we do in Table 1.3. The first column of Table 1.7 shows that, when ignoring worker gender, we see positive returns to negotiation for Typing-workers and column 3 shows that we observe no return to negotiation for Slider-workers. This matches what we observed in our full sample regressions. Specifically, the coefficient on Entry (a dummy variable that equals one in periods that a worker enters negotiation) is 21.1. This reveals a positive net return from negotiation (probability different from \$0.05,  $p=0.033$ ). Thus, as with our full sample, we confirm that negotiation only benefits the workers assigned to the high-productivity task.

Next, we consider how our results change when we separately control for gender and the interaction between negotiation and gender. In the second column of Table 1.7, we show positive returns to negotiation for Typing-workers of 38.1 cents (probability different from \$0.05,  $p=0.001$ ). Interestingly, however, these returns are only experienced by male Typing-workers. The interaction term Entry\*Female is negative, highly significant, and nearly the same as the magnitude for the coefficient on Entry. Thus, the positive return to negotiation is fully cancelled out for female Typing-workers. We observe no such gender differences in returns to negotiation when we look at the Slider task (Column 4 of Table 1.7). These results partially confirm our hypothesis that only male workers benefit from negotiation. Specifically, we find that only male workers assigned to the Typing-task benefit from negotiation. Female Typing-workers and Slider-workers of both genders do not benefit. The results also are consistent with our earlier findings – negotiation does not benefit workers on the low productivity task.

Given that we observe male Typing-task workers earning a positive return, but not female Typing-task workers, we test whether this leads to gender earnings gap conditional on negotiation. To do this, we test whether Female+Entry\*Female is different from zero. We find that the earnings gap conditional on negotiation is 29.0 cents and thus reject the null that this gap equals zero ( $p=0.074$ ). As a result, we conclude that under negotiation, we observe a gender earnings gap among high-productivity workers that negotiate.

The finding that only male, high-productivity task workers benefit from negotiation is consistent with our finding that male-typing task workers experience a reduced earnings advantage following a Negotiation-Ban. As further evidence that it is this differential return to entry driving the gender differences in compensation differences across treatments, we regress Typing-worker earnings on output and negotiation entry pooling together both treatments. We add an indicator variable for the Negotiation-Ban treatment, an indicator variable for negotiation entry by workers,

Table 1.7: Per Period Worker Earnings: Negotiation Treatment, Mixed Gender Groups

	(1)	(2)	(3)	(4)
	Typing-Worker	Typing-Worker	Slider-Worker	Slider-Worker
# Letter-pairs	0.0386*** (0.00243)	0.0385*** (0.00258)	0.0120*** (0.00274)	0.0118*** (0.00293)
# Sliders	0.0273*** (0.00866)	0.0322*** (0.00815)	0.0249*** (0.00905)	0.0223** (0.00885)
Entry	0.211*** (0.0755)	0.381*** (0.0962)	0.0904 (0.0906)	0.112 (0.134)
Female		0.00244 (0.157)		-0.122 (0.178)
Entry X Female		-0.395*** (0.147)		-0.0509 (0.172)
Constant	-0.617** (0.271)	-0.727** (0.303)	0.415 (0.310)	0.555 (0.342)
N	144	144	144	144

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Observations are restricted to mixed-gender groups only and first four rounds of data. Standard errors are clustered at the group level and are shown in parentheses. Results are from random effects regression on per-period worker earnings. Entry is a binary variable that equals one if the worker enters negotiation in the given period. All regressions include period fixed effects.

and interact both with worker gender. Doing this allows us to test whether we observe differences in earnings across treatments (both overall and by worker gender) after controlling for gender differences in returns to negotiation. As shown in Table 1.8, neither Negotiation-Ban, Female, nor their interaction are significant. In contrast, Entry is positive and significant and Entry x Female is negative, significant, and of a similar magnitude to the coefficient on Entry. Thus, gender differences in the effects of a Negotiation-Ban on compensation inequality are due to gender differences in returns to negotiation.

To summarize, we find that only male workers assigned to the high productivity task experience positive returns to negotiation. Under a Negotiation-Ban policy, worker output does not decrease, and the relative earnings advantage of male typing-task workers does decrease. As a result, a ban leads to a reduction in the gender gap in compensation.

## 1.5 DISCUSSION

In this section, we ask why it is that only high-productivity males benefit from negotiation. When considering the role negotiation plays in determining compensation, workers may use negotiation to try and sway manager beliefs about their relative task difficulty, they may try to use negotiation to win the manager over via likeability, or they may try to threaten the manager with reduced future performance. If negotiation tactics have differing levels of effectiveness and workers differ by gender in their method of negotiating, this could generate the observed differences in returns to negotiation. In this section, we separately consider each of these possible explanations.

### 1.5.1 Does Negotiation Affect Manager Beliefs?

We first ask whether manager beliefs differ by treatment and worker-gender composition. If negotiation affects manager beliefs, then we should observe this when comparing manager survey responses across treatments. Similarly, if male Typing-workers are more effective at swaying manager beliefs, we should see manager beliefs in the Negotiation treatment differing by worker-gender composition.

We consider two different measures for manager beliefs. The first measure Letter-pairs Needed is the manager's reported number of letter-pairs they believe are equal in effort to 15 sliders. This measures manager beliefs about relative difficulty. Our second measure for manager beliefs, Fair Allocation, is the manager reported share they believe is fair to allocate to the Typing-worker after observing output of 25 sliders and 75 letter-pairs. This question is intended to measure manager perceptions of fairness.

Looking across treatments (Table 1.9, panel a), and across worker-gender composition by treatment (Table 1.9, panels b & c), we observe no significant differences in manager beliefs of relative difficulty or of fairness. Thus, it appears unlikely that the observed returns to negotiation are due to workers affecting manager beliefs. Not surprisingly, when looking at worker beliefs, we do observe significant differences by assigned role. Slider-workers believe 60.35

Table 1.8: Typing Worker Compensation Pooling over  
Both treatments: Mixed Gender Pairs

	(1)	(2)
	Typing-Worker	Slider-Worker
# Letter-pairs	0.0354*** (0.00287)	0.0149*** (0.00296)
# Sliders	0.0249*** (0.00703)	0.0280*** (0.00733)
Entry	0.388*** (0.0985)	0.109 (0.125)
Female	0.0251 (0.157)	-0.109 (0.171)
Entry X Female	-0.390*** (0.146)	-0.0365 (0.165)
Negotiation-Ban	-0.168 (0.128)	0.0699 (0.190)
Negotiation-Ban X Female	0.144 (0.224)	0.288 (0.236)
Constant	-0.300 (0.300)	0.133 (0.326)
N	268	268

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Observations are restricted to mixed-gender groups only and first four rounds of data. Standard errors are clustered at the group level and are shown in parentheses. Results are from random effects regression on per-period worker earnings. Entry is a binary variable that equals one if the worker enters negotiation in the given period. Negotiation-Ban is a binary variable that equals one in the Negotiation-Ban treatment. All regressions include period fixed effects.

letter-pairs are equivalent to 15 sliders whereas Typing-workers believe only 48.15 letter-pairs are equivalent to 15 sliders. Similarly, Slider-workers believe a fair allocation for the Typing-worker given an output of 25 sliders and 75 letter-pairs is 51% whereas Typing-workers believe a fair allocation is 61%. Both differences are significant at the 1% level and show workers beliefs are somewhat self-serving.

To further test how manager beliefs about relative task difficulty affect their compensation decisions in the negotiation treatment, we turn to regression analysis. Using random effects regressions with standard errors clustered at the group level, we examine how manager beliefs affect Typing-worker earnings in the first four periods. Table 1.10 shows that beliefs do matter for compensation. A 10 unit increase in Letter-pairs Needed is associated with around a \$0.06 to \$0.12 decrease in the amount they give to the Typing-worker depending on the specification. To further help interpret this, suppose a manager believes the Typing task is twice as difficult as the Slider task and thus report that 30 letter-pairs are equal to 15 sliders in effort, they will give around \$0.9 to \$0.18 less to the Typing-worker in each period than a manager that believes the two tasks are equally difficult. Similarly, beliefs about fairness also matter. A 10-percentage point increase in Fair is associated with around a 0.23 to 0.26 increase in the amount allocated to the Typing-worker. Furthermore, both beliefs appear to matter even when controlling simultaneously for both beliefs (column 3).

Interestingly, while beliefs matter for manager compensation decisions, including manager beliefs as controls does not significantly affect estimated returns to negotiation on the Typing task. Table 1.10 shows that, after controlling for beliefs, the point-estimate for Entry is positive and significant when considering the first four periods of data. This matches what we observed in Table 1.3. Thus, it seems unlikely that the mechanism leading to differences in returns to negotiation could be due to Typing-workers being more effective at swaying manager beliefs than Slider-workers or Typing-workers using this method in their negotiation more often than other workers.

Looking by gender and restricting our attention to only mixed-gender groups, we observe a similar pattern. Table 1.11 shows that a 10 unit increase in Letter-pairs Needed is associated with between a \$0.06 to \$0.09 cent decrease (depending on the specification) in the amount they give to the Typing-worker. Similarly, a 10-percentage point increase in Fair is associated with between a \$0.17 to \$0.20 increase in the amount they give to the Typing-worker. Controlling for manager beliefs does not eliminate the gender difference we observe in returns to entering negotiation. Thus, it also seems unlikely that male Typing-workers are more effective at swaying manager beliefs than female Typing-workers.

### 1.5.2 Differences in Chat Style

In this subsection, we ask whether differences in chat style attenuate the gender gap in returns to negotiation. To identify chat styles, we hired four undergraduate research assistants to independently code all negotiations. We specified ten different categories/characteristics that we felt represented the different behaviors that might arise in the negotiations. These ten categories are listed in Table 1.12. Coders could select as many or as few categories that they felt

Table 1.9: Stated Beliefs

Panel a. Manager Beliefs by Treatment	Ban	Negotiation	Difference (Standard Error)
Letter-pairs Needed	52.76	49.40	3.36 (3.91)
Fair Split	0.57	0.57	-0.00 (0.03)
N	42	42	
Panel b. Manager Beliefs by Gender- Composition: Negotiation Treatment	Male Typing	Female Typing	Difference (Standard Error)
Letter-pairs Needed	50.00	48.24	1.76 (6.09)
Fair Split	0.56	0.59	-0.03 (0.04)
N	19	17	
Panel c. Manager Beliefs by Gender- Composition: Ban Treatment	Male Typing	Female Typing	Difference (Standard Error)
Letter-pairs Needed	55.29	50.21	5.08 (6.55)
Fair Split	0.53	0.58	-0.05 (0.05)
N	17	14	
Panel d. Worker Beliefs by Role: Both Treatments	Slider- Worker	Typing- Worker	Difference (Standard Error)
Letter-pairs Needed	60.35	48.15	12.19*** (3.05)
Fair Split	0.51	0.61	-0.10*** (0.02)
N	84	84	

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Letter-pairs Needed is the individual's response to a question asking how many Letter-pairs are equal 15 Sliders worth of effort. Fair Split is the Individual's response to a question asking what allocation would be fair for the Typing-worker to receive given an output of 25 Sliders and 75 Letter-Pairs.

Table 1.10: Effect of Manager Beliefs on Typing-Worker Pay

	(1)	(2)	(3)
# Letter-pairs	0.0385*** (0.00261)	0.0378*** (0.00256)	0.0384*** (0.00240)
# Sliders	0.0143** (0.00719)	0.0130** (0.00649)	0.0108 (0.00662)
Entry	0.232*** (0.0660)	0.235*** (0.0668)	0.234*** (0.0679)
Letter-pairs Needed	-0.0121*** (0.00350)		-0.00579* (0.00321)
Fair Split		2.568*** (0.599)	2.266*** (0.595)
Constant	0.260 (0.340)	-1.715*** (0.443)	-1.257** (0.526)
N	168	168	168

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Standard errors are clustered at the group level and are shown in parentheses. Observations are restricted to the first four periods. Results are from random effects regressions of money given to the Typing-worker within a period in the Negotiation treatment. Entry is a binary variable that equals one if the worker enters negotiation in the given period. Letter-pairs Needed is the manager's response to a survey question asking how many Letter-pairs are needed to equal 15 Sliders worth of effort. Fair is the manager's response to a survey question asking what allocation would be fair for the Typing-worker to receive given an output of 25 Sliders and 75 Letter-Pairs. All regressions include period fixed effects.

Table 1.11: Manager Beliefs on Typing-Worker Pay: Mixed-Sex

	(1)	(2)	(3)
# Letter-pairs	0.0400*** (0.00278)	0.0383*** (0.00244)	0.0393*** (0.00240)
# Sliders	0.0259*** (0.00926)	0.0249*** (0.00827)	0.0214** (0.00915)
Entry	0.390*** (0.0968)	0.352*** (0.0971)	0.360*** (0.0982)
Female Typing-worker	0.0118 (0.147)	-0.0563 (0.119)	-0.0428 (0.118)
Entry X Female Typing-worker	-0.419*** (0.148)	-0.329** (0.146)	-0.348** (0.148)
Letter-pairs Needed	-0.00945*** (0.00303)		-0.00581* (0.00306)
Fair Split		2.004*** (0.654)	1.740*** (0.630)
Constant	-0.244 (0.377)	-1.680*** (0.396)	-1.249*** (0.453)
N	144	144	144

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Observations are restricted to mixed-gender groups only and the first four periods. Standard errors are clustered at the group level and are shown in parentheses. Results are from random effects regressions of money given to the Typing-worker within a period in the Negotiation treatment. Entry is a binary variable that equals one if the worker enters negotiation in the given period. Letter-pairs Needed is the manager's response to a survey question asking how many Letter-pairs are needed to equal 15 Sliders worth of effort. Fair is the manager's response to a survey question asking what allocation would be fair for the Typing-worker to receive given an output of 25 Sliders and 75 Letter-Pairs. All regressions include period fixed effects.



applied to a given negotiation. Negotiations were coded at the worker-manager level (i.e., coding for a given worker was based on all chat interaction between that worker and the manager over the five periods of the experiment).

The coders were unfamiliar with the research questions. Coders were given a summary of the experimental instructions and a list of the ten categories with examples.<sup>17</sup> Coders viewed the full set of messages sent between each worker-manager pair, the worker's role, and the period in which each message was sent. Given that there is not perfect agreement among coders, we categorize a worker as exhibiting a characteristic if at least three of the four coders characterized them as such.

Table 1.12 provides a brief description of each category. Overall, we observe different chat characteristics exhibited by Slider and Typing-workers in a manner that capture some of the key tensions we our design sought to create; 35% of Typing-workers are coded as aggressive while only 9% of Slider-workers are. Workers are much more likely to argue for the fairness norm that works in their favor. Slider-workers are much more likely than Typing-workers to argue for merit-based fairness norms (85% versus 6%), whereas Typing-workers are much more likely than Slider-workers to argue for output-based fairness norms (76% versus 0%). Only Typing-workers are coded as other regarding (12% versus 0%); however, this is not surprising given that Typing-workers are in a position where it is easier to be other-regarding. From this, we conclude that workers are using the chat feature to negotiate and that our coding of chat data makes sense.<sup>18</sup>

### 1.5.3 Is the Gender Gap in Returns to Negotiation Attenuated by Differences in Negotiation Style?

In this subsection, we restrict our attention to mixed-gender pairs to analyze whether we observe differences in negotiation characteristics and/or differences in returns to negotiation characteristics by worker gender and task. Table 1.13 provides a summary of the observed characteristics among workers that entered negotiation split by worker gender and pooling over the two tasks. We find several differences in negotiation characteristics by worker gender. As we expected, we observe that a much smaller proportion of female workers use threats when negotiating. Only 4% of female workers use a threat when negotiating; however, 24% of male workers use a threat when negotiating. Males are more likely to argue for an even split (17% versus 4%); however, female workers are more likely to make statements about merit (57% versus 31%). Finally, we observe no female workers using humor, but we observe 28% of male workers using humor.

Given the large number of differences observed in negotiation characteristics between male and female workers, we ask whether these differences attenuate the gender gap in returns to negotiation entry. To do this, in Table 1.14, we modify our baseline regression specification (repeated in column 1 of this table) to include observed characteristics.

---

<sup>17</sup>We exclude the category Compromise from all analysis. When debriefing coders after they completed all tasks, it become clear that this category was being viewed by them as a proxy for whether they observed a successful negotiation rather than representing a worker compromising on their initial request.

<sup>18</sup>Following Cooper and Kühn (2014), for each category we also calculated Cohen's Kappa (Cohen 1960) to provide a measure of intra-coder agreement. Using the guidelines from Landis and Koch (1977), we note that the majority of our categories have moderate to substantial agreement between coders.

Table 1.12: Frequency of Characteristics Among Workers that Negotiated

	Pooled	Slider Worker	Typing Worker	Difference by Task (S.E.)
<i>Aggressive - The worker is assertive, direct, entitled, or aggressive</i>	0.23	0.09	0.35	-0.26** (0.10)
<i>Threat - The worker directly or indirectly threatens to reduce their output in subsequent rounds</i>	0.12	0.13	0.12	0.01 (0.08)
<i>Deferential - The worker talks in a non-confrontational manner, uses caveats, is apologetic, or is uncertain</i>	0.21	0.19	0.24	-0.05 (0.10)
<i>Even Split - The worker explicitly says joint worker earnings should be split equally</i>	0.09	0.13	0.06	0.07 (0.07)
<i>Merit - The worker proposes that the relative task difficulty should be taken into account</i>	0.44	0.84	0.06	0.78*** (0.08)
<i>Output - The worker proposes that earnings should be based solely on output</i>	0.39	0.00	0.76	-0.76*** (0.08)
<i>Humor - The worker uses humor when negotiating with the manager</i>	0.12	0.19	0.06	0.13 (0.08)
<i>Other-Regarding - The worker explicitly suggests the other worker's well-being should be taken into account</i>	0.06	0.00	0.12	-0.12** (0.06)
<i>Off Topic - The majority of the conversation is not related to the negotiations or the task.</i>	0.06	0.06	0.06	0.00 (0.06)
N (Number of Workers)	66	32	34	

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Standard Errors are in parentheses. Aggressive, Threat, Deferential, Even Split, Merit, Output, Humor, Other Regarding, and Off Topic are binary variables that each equal one if the worker was coded as exhibiting that characteristic by at least 3 of the 4 coders.

Table 1.13: Frequency of Characteristics Among Workers that Negotiated: Mixed Gender

	Pooled	Male Worker	Female Worker	Difference by Gender (S.E.)
<i>Agressive - The worker is assertive, direct, entitled, or aggressive</i>	0.23	0.21	0.25	-0.04 (0.11)
<i>Threat -The worker directly or indirectly threatens to reduce their output in subsequent rounds</i>	0.14	0.24	0.04	0.21** (0.09)
<i>Deferential - The worker talks in a non-confrontational manner; uses caveats, is apologetic, or is uncertain</i>	0.25	0.28	0.21	0.06 (0.12)
<i>Even Split - The worker explicitly says joint worker earnings should be split equally</i>	0.11	0.17	0.04	0.13* (0.08)
<i>Merit - The worker proposes that the relative task difficulty should be taken into account</i>	0.44	0.31	0.57	-0.26** (0.13)
<i>Output - The worker proposes that earnings should be based solely on output</i>	0.39	0.41	0.36	0.06 (0.13)
<i>Humor - The worker uses humor when negotiating with the manager</i>	0.14	0.28	0.00	0.28*** (0.09)
<i>Other-Regarding-The worker explicitly suggests the other worker's well-being should be taken into account</i>	0.07	0.07	0.07	-0.002 (0.07)
<i>Off Topic - The majority of the conversation is not related to the negotiations or the task.</i>	0.035	0.034	0.035	-0.001 (0.05)
N (Number of Workers)	57	29	28	

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Standard Errors are in parentheses. Observations are restricted to mixed-gender groups only. Aggressive, Threat, Deferential, Even Split, Merit, Output, and Humor are binary variables that each equal one if the worker was coded as exhibiting that characteristic by at least 3 of the 4 coders.

Table 1.14: Typing-Worker Pay Controlling  
for Chat Characteristics: Mixed Gender

	(1)	(2)
# Letter-Pairs	0.0385*** (0.00258)	0.0394*** (0.00288)
Sliders	0.0322*** (0.00815)	0.0363*** (0.00765)
Entry	0.381*** (0.0962)	0.206* (0.111)
Female	0.00244 (0.157)	0.0288 (0.159)
Entry X Female	-0.395*** (0.147)	-0.264** (0.122)
Threat		0.693** (0.314)
Agressive		0.0825 (0.143)
Humor		0.0224 (0.244)
Deferential		0.0499 (0.134)
Output		0.00752 (0.139)
Other-Regarding		-0.163 (0.269)
Constant	-0.727** (0.303)	-0.922*** (0.332)
N	144	144

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Observations are restricted to first four periods and mixed-gender groups only. Standard errors are clustered at the group level and are shown in parentheses. Results are from random effects regression on per-period typing worker earnings. Column 1 replicates the baseline results for reference. All regressions include period fixed effects.

We restrict our attention to characteristics in which at least 3 negotiating workers are coded as exhibiting the characteristic.<sup>19</sup> Each characteristic is an indicator variable that takes on a value of 1 for a given worker-period if the worker was coded as exhibiting the characteristic and entered negotiation in that period. As shown in column 2, controlling for negotiation characteristics return to negotiation entry for male Typing-workers by almost one-half from 38 cents to 20 cents. The attribute that provides the largest return to negotiation is the use of threats.<sup>20</sup> Given that we observe only male workers using threats, one may suspect that this difference in negotiation style is what is driving the gender gap in returns to negotiation. While we do observe that controlling for the negotiation attributes causes the gender gap in returns to negotiation entry to decrease in size by a bit less than one-half, the gap is still sizeable and statistically significant. This suggests that gender differences in negotiation style (such as female workers lack of using threats) only partially attenuates the gender gap in returns to negotiation.

## 1.6 CONCLUSION

In recent years, some corporations have enacted Negotiation-Ban policies to combat gender differences in compensation. The reasoning for enacting such policies stems from the commonly held belief that gender differences in negotiation are one of the primary factors contributing to compensation differences. Evidence of implicit bias, however, suggests that differential treatment also can arise when management alone selects worker compensation. Previous literature on negotiation largely focuses on studying either gender differences in entrance into negotiation and gender differences in returns to negotiation.

This study differentiates itself from previous literature by presenting a first examination of a Negotiation-Ban counterfactual. We investigate the manager-selected compensation that results when workers can and cannot negotiate their salary in an environment where workers perform either high or low productivity tasks. Differentially productive tasks make room for subjective assessment and for concerns about future productivity and increase the possibility for implicit biases to affect compensation decisions. When negotiations are permitted we find that the more productive task is better compensated, and that the more productive task compensation is greater for men. Thus, a gender wage gap arises on the high productivity task. We demonstrate that this gap is not fully attenuated by gender differences in negotiation style.

Intriguingly a Negotiation-Ban reduces the relative pay advantage of the high-productivity task, does so more for men than for women, and reduces overall earnings inequality among workers. Thus, consistent with the recent push to ban salary negotiations, we find that the gender gap in compensation is eliminated both on the low and high productivity task under a Negotiation-Ban policy.

<sup>19</sup>Removing this restriction does not change the qualitative interpretations of our results. The exception is that the estimated coefficient for Threat remains becomes less precise due to multicollinearity.

<sup>20</sup>We note that using threats as a Slider-task worker is not beneficial. Table 1.14 shows that Slider-task workers using threats instead experience a negative return to negotiation. This is likely due to the lower bargaining power held by Slider-task workers.

## 2.0 GENDER DIFFERENCES IN EXECUTIVE DEPARTURE

### 2.1 INTRODUCTION

Over the past fifty years, the U.S. has seen substantial progress in narrowing the gap in the representation and compensation of women in the labor market (Bailey and DiPrete, 2016). However, a pervasive gender gap still persists at the top of the income distribution (Guvenen et al., 2014). One area that is especially conspicuous is the gap among top executives at publicly traded firms. A large body of research documents that women are generally paid less than their male counterparts and are underrepresented in top management positions.<sup>1</sup> Many explanations have been put forth to explain the underrepresentation of women at this level. Explanations include gender differences in preferences (e.g., fertility and family ties) and ability; discrimination and organizational barriers such as masculine work culture and exclusion from social networks (Kark and Eagly, 2010); and, more recently, it has been argued that behavioral differences such as differences in willingness to compete and confidence (Niederle and Vesterlund, 2007) as well as differences in negotiation (Babcock and Laschever, 2003) contribute to the gap.

While these are certainly important components to the underrepresentation of women in top management positions, they tend to focus only on differences in entry. It is not enough, however, to facilitate women's entry; we must also focus on their retention. A smaller set of studies focuses on exactly that (Becker-Blease et al., 2016, 2010; Gayle et al., 2012; Guest, 2016). Overall, the results consistently demonstrate that female executives exhibit higher departure rates than male executives. While the existing papers explore a variety of factors that may be correlated with the gender gap in departure, none of the studies provide evidence that allows us to rule out the simplest explanation for this observed difference: gender differences in ability.

This chapter examines whether the observed gender differences in executive departure can be explained solely by gender differences in ability. To address this, I construct an instrument for firm performance that strips away the ability component, use the instrument to predict departure rates, and examine whether the gender gap in departure is correlated with the instrument.

This instrument is constructed using industry mean performance (with the own-firm removed) to instrument for firm performance, as is done in the relative performance evaluation literature (Fee et al., 2015; Guay et al., 2014; Jenter

---

<sup>1</sup>See for instance Kark and Eagly (2010) or Blau and Kahn (2016) for review.

and Kanaan, 2006, 2015), and is used to estimate the effect of firm performance on gender differences in departure probability. Specifically, I use 2SLS to instrument for negative firm performance (as defined by having a negative annual return) and the interaction between negative firm performance and executive gender. The excluded instruments I use are negative firm-removed industry performance and its interaction with gender. Given that these instruments are exogenous from executive ability within a firm, they allow for identification of whether gender differences from departure are coming from factors outside of ability.

Using matched employer-employee data of S&P 1500 executives, and focusing primarily on industry contractions, the results show that following a contraction, the gender gap in executive departure rates increases by around 5 percentage points. Further, there is no effect on the departure rate for male executives, suggesting this increase in the departure rate gap is driven solely by an increase in female departure rates. Given a base departure rate for female executives of 13.6%, and a gender gap, this amounts to approximately a 35% increase in the departure rate for female executive following exogenous contractions in firm performance.

Because the gender difference in departure rates is affected by exogenous changes in firm performance, this shows that the gender gap in executive departure cannot be explained by women's under performance alone. As such, it is not sufficient for policies focused on closing the gender gap in departures to simply target improving women's leadership skills. Instead, such policies should target the underlying mechanisms that are driving the gender gap.

Given this result, the remainder of the chapter explores alternative mechanisms and provides compelling evidence to rule them out. Specifically examined are the following: gender differences in fertility, early retirement, external hiring, the glass cliff (whereby women may be in more tenuous positions to begin with), and female start-ups. Supported with findings from the literature, in addition to the data, it is clear that none of these potential mechanisms is likely to be driving the observed treatment effect.

One may ask, if none of the above mechanisms are driving the effect, what could remain? One possibility that remains consistent with the results in this study is the conjecture of gender differences in misplaced blame.<sup>2</sup> While I am unable to identify this mechanism with the existing data, this mechanism would also align with existing literature that shows female executives' pay decreases more than that for male executives following negative firm performance, whereas the opposite pattern emerges following positive firm performance (Selody, 2010; Albanesi et al., 2015).

The rest of this chapter proceeds as follows: in section 2.2, I provide an overview of the executive labor market, and discuss related literature on executive departures. In section 2.3, I discuss the data used in this study and how variables are coded. In section 2.4, I discuss my methodology. Section 2.5 presents the main results, while section 2.6 explores potential first-order mechanisms for the observed treatment effect. In section 2.7, I discuss blame as an alternative explanation for gender differences in departures. Finally, section 2.8 concludes.

---

<sup>2</sup>Examples of misplaced blame include attribution bias in which the board misattributes poor firm performance from industry shocks as being due to ability when the executive is female. Another possibility could be self-attribution bias in which female executives more frequently misattribute failures that are outside their control to internal factors.

## 2.2 RELATED LITERATURE

### 2.2.1 The Executive Labor Market

Before discussing the details of the gender gap in executive departure rates, we must first define what constitutes a top executive position. Top executives are the highest ranked executives in a company and belong to the top management team (TMT) within the organization. As defined in (Castanias and Helfat, 1991), and in (Hambrick and Mason, 1984), a TMT consists of the CEO, and the senior executives that report directly to the CEO. These positions are typically at or above the level of vice president, including titles such as CFO, COO, etc.

According to Murphy (1999), the majority of top executive positions have pay packages with four distinct components in common. These components are a minimum salary, a performance-based yearly bonus, the option to purchase stocks at a discounted rate – which is particularly prevalent among U.S. executives, often comprising more of an individual executive’s total compensation than their minimum salary – and an incentivization scheme for encouraging long-term performance. Furthermore, unlike typical at-will employees, who comprise the low- and mid-level ranks of a company, top executives will usually have formalized contracts, lasting a certain number of years and stipulating the details of the above components, as well as potential severance packages.

Given these characteristics, particularly the existence of a formalized contract with severance, these positions will prove much more resilient to downsizing in times of economic downturn. According to Cameron et al. (1993), downsizing may encompass eliminating individuals or work assignments, organizational restructuring, partial dissolution, and full dissolution of the company. As discussed in Krishnan and Park (1998), when downsizing occurs, top management teams have the tendency to replace executives perceived as underperforming, rather than simply eliminating their positions entirely. As such, the reasons behind a top executive’s dismissal are more likely to be focused on the perception of the difference in their ability versus the abilities of their peers than they are to be focused on the simple redundancy of her position. This is perhaps self-evident, given that TMT positions are necessary for the operation of a firm regardless of size. As such, they cannot easily be made redundant.

### 2.2.2 Executive Departure

A number of factors have been shown to affect executive departure. Most obviously, it has been documented that executive departure is linked to the executive’s performance – e.g., Homroy (2015), Guay et al. (2014). Involuntary departures have also been attributed to the adaptability of executives to long term industry shocks, such as changes in the industry concentration, with generalists having lower involuntary departure following such shocks than other executives (Guay et al., 2014).

Gender is also contributing factor to executive departure, with female executives exhibiting both higher voluntary and involuntary departure rates (Becker-Blease et al., 2010, 2016; Gayle et al., 2012; Guest, 2016). The gender gap



in departure rates tends to be largest in smaller firms, firms with smaller boards, and firms with fewer independent directors (Becker-Blease et al., 2010). However, the effect is not necessarily limited to such firms. In fact, the gender difference in departure rates has significant effects on the representation and compensation of women in executive positions (Gayle et al., 2012; Becker-Blease et al., 2016). This effect is great enough that Gayle et al. finds that the commonly documented gender difference in executive pay can be largely explained by these differences in executive departure.

Concordantly, researchers have been interested in exploring whether improving female representation among board members may help reduce this departure rate gap. It is unclear, however, whether increases in female leadership helps to attenuate the gender gap in departure rates. While some evidence suggests increased female representation on the board of directors leads to a decrease in the gender gap in departures (Becker-Blease et al., 2010, 2016), conflicting findings in Guest (2016) suggest that improved female leadership at the CEO or board level does not reduce the gender gap in departure. While it is important to identify factors that may help reduce the gender gap in departures, a first order question is whether this gap in departures can be rationalized by differences in ability alone. If the answer is ‘yes,’ then effecting organizational change to reduce female departure rates is covering the hole rather than repairing it. Instead, research should focus on identifying factors such as systemic and institutionalized pressures which may be impacting women’s ability.

To my knowledge, this is the first study to explore whether the observed gender differences in departure can be fully explained by gender differences in ability. Along with addressing this, my study contributes by investigating other, previously unexplored mechanisms that may contribute to gender differences in departure.

## 2.3 DATA

### 2.3.1 Data Overview

This study uses data from Standard and Poor’s Execucomp database of the S&P 500, S&P MidCap 400, and S&P SmallCap 600 companies. Execucomp is an annual dataset that is compiled by S&P Capital IQ using each included company’s annual DEF14a proxy statement filed with the SEC. Execucomp contains compensation data on up to fifteen of the top executives within each firm in a given year, as determined by the executives’ annual total compensation.<sup>3</sup> Commonly seen top executive officer titles within the sample include CEO, CFO, COO, President, Executive VP, and chairman of the board. This dataset contains unique executive and firm identifiers, allowing executives to be tracked if they move from one Execucomp company to another. In addition to providing information on annual

---

<sup>3</sup>DEF 14a SEC proxy statements require the firm to list the CEO, CFO, and top 3 non-CEO non-CFO executive officers (determined by total compensations) that were employed as of the end of the fiscal year. Firms also must include up to 2 non-CEO, non-CFO executives that would have been listed had they been employed by the end of the fiscal year. Some firms choose to report more executives than the number required by the SEC.

executive compensation, Execucomp provides details on executive gender, title, and age. I merge Execucomp with Compustat data for more detailed firm level information and with data from the Center for Research in Security Prices data for annual stock returns values, my primary measure of firm performance.

I restrict the sample to only include data from 1998 to 2015 due to the low number of female executives prior to 1998 and incomplete data from 2016 forward. As in [Selody \(2010\)](#), observations with missing data or negative observations for total compensation are dropped from the sample. Also restricted from the sample are executives with missing age data, a reported current age below 18, missing gender data, firms that have missing data on annual stock returns, firms with missing data on the number of employees, and all firms whose industry is classified as “other” per the Fama-French 48 industry classifications ([Fama and French, 1997](#)). Finally, in order to identify tenure for the executives in the sample, executives that appear as a top executive within the firm prior to 1997 are excluded. Tenure is measured as the number of years in which an executive appears among the top executives at the firm, as defined by the firms’ SEC proxy statement.<sup>4</sup> Doing this allows for the identification of tenure for all executives in the sample without incurring any left censoring of this variable. One caveat of this approach is that it creates a measure of tenure that does not identify an executive’s overall tenure as an employee within the firm, but rather their tenure as a top executive within the firm.

The final sample consists of 84,998 executive-firm-year observations. Within this sample, there exist 2,278 unique firms, 19,626 unique executive-firm pairs, and 18,504 unique executives. Thus, only a small portion of executives within the sample appear at multiple firms. Focusing on female executives, there are 6,846 executive firm-year observations with 1,805 unique female executive-firm pairs and 1,717 unique female executives. Thus, female executives comprise approximately 8.05% of the executive-firm-year observations, 9.2% of the executive-firm pairs, and 9.3% of the unique executive observations. The share of female executives observed in the sample is higher than the 4.5% observed in [Becker-Blease et al. \(2010\)](#); however, representation of women in top executive positions has increased over time.<sup>5</sup> Looking instead at studies using slightly more recent data, we can see that 9% is quite comparable to that observed in other samples using a more similar date range. For instance, [Selody \(2010\)](#) finds between 7 to 8% of top executives were female between 2004 and 2008.

### 2.3.2 Identifying Executive Departures

While Execucomp includes some departure dates in their data, the reporting of this information was never mandatory. Furthermore, Execucomp stopped collecting this information entirely after 2009. As a result, departure date information in Execucomp is largely incomplete prior to 2009 and nearly non-existent following 2009 ([Stefanescu et al.](#),

---

<sup>4</sup>Execucomp provides start date information for a subset of executives in the sample; however, this variable is often not reported. As a robustness check, I repeat the primary analysis including executives with earlier start dates. The results are largely unchanged by relaxing this restriction. See the robustness checks subsection in section 2.5.

<sup>5</sup>Focusing on my sample, only 4.6% of executives were female in 1998 compared to over 8% by 2008.

2015).<sup>6</sup> Thus, I construct an alternative measure of departure using the method described in [Guay et al. \(2014\)](#). An executive is flagged as having departed firm  $f$  in year  $t$  (DEPART) if the following two conditions are met: they are not listed as an executive at firm  $f$  for years  $t + 1$  and  $t + 2$ ; and the executive did not lose their job as a result of the firm shutting down.

The primary goal of this study, however, is to identify which factors may be contributing to differences in involuntary departure rates between male and female executives. Unfortunately, identifying involuntary departures poses its own set of challenges. Reasons behind executive departures are largely kept private, and — except under extreme circumstances — all departure events are generally referred to as resignations, without indication as to whether they were voluntary. Methods of coding involuntary departures in previous studies range from treating all departures as voluntary — unless they can be identified with a specific news event — to the other extreme: coding all departures as involuntary, unless given strong evidence to suggest otherwise. Given the sparse nature of news reports on executive resignation events, many researchers that rely on using news reports employ some version of the [Parrino \(1997\)](#) algorithm that dictates how to handle departure events for which no news article exists. One shortcoming of relying on news articles to categorize turnover is that news reporters may be subject to biases, as argued in both [Fee et al. \(2015\)](#) and [Jenter and Lewellen \(2010\)](#). In particular, [Fee et al. \(2015\)](#) argues that reports may be more likely to suggest a resignation was forced when the firm performance is negative, regardless of the actual reason for departure.

Potential biases on behalf of news reporters can become even more problematic when introducing the additional dimension of gender. For instance, a media analysis focused on news coverage of CEOs during times of transition found that the framing of stories is much more likely to attribute blame to the CEO when she is female ([The Rockefeller Foundation and Global Strategy Group, 2016](#)). In this instance, we have to be concerned not only with reporters having a greater inclination to report forced resignation when firm outcomes are negative; we must also be concerned with how any inherent gender biases on behalf of the reporters shade their news reports.

In light of potential biases that can be introduced through relying on media reports, I instead rely on a modified method from [Fee et al. \(2015\)](#), in which an executive is flagged as having “involuntarily” departed (INVOLUNTARY) if they satisfy the conditions for having departed (DEPART) and they do not reappear as a top executive at another Execucomp firm within two years of departure.<sup>7</sup> This measure attempts to rule out executive turnover from voluntary job changes by executives (e.g., lateral or upwards movements). This measure is not able to rule out executives that voluntarily leave the labor market. As such, in section 2.6 of this chapter, I explore the possibilities of fertility or early retirement. As an additional robustness check, I also construct a measure, INVOLUNTARY2, that includes the same restrictions as INVOLUNTARY, but does not mark as involuntarily departed the small set of executives that Execucomp has identified as having departed due to being retired or deceased.<sup>8</sup>

<sup>6</sup>Execucomp also provides reasons for departure on the executives they flag as having departed. These categories are quite coarse; they are limited to “deceased,” “retired,” “resigned,” and “unknown.”

<sup>7</sup>The differences between using a one year and two year restriction for reappearance are small (refer to table 2.1). Using 2 years is a slightly more conservative definition, as it allows for any friction executives may experience in job transition and reduces the likelihood that executives taking leave for fertility reasons will be marked as departed.

<sup>8</sup>I choose to use INVOLUNTARY instead of INVOLUNTARY2 as my main measure of departure, given the possibility for bias if the missing

### 2.3.3 Firm Performance

In order to identify gender differences in executive departure that are exogenous from executive ability, I construct an instrument for firm performance. For a given firm  $f$ , the instrument is equal to the mean performance of all firms in firm  $f$ 's industry with firm  $j$  removed from the average. Thus, this instrument represents the average industry performance for firm  $f$ 's peers defined at the industry level. Industries are defined using the Fama-French 48 industry classification (Fama and French, 1997).

Firm performance is measured using the firm's annual stock return as retrieved from CRSP. Stock returns are one of the most commonly used performance measures in studies of executive compensation and departure – e.g., Becker-Blease et al. (2010), Fee et al. (2015), Guest (2016), Jenter and Kanaan (2015). Unlike accounting-based measures, stock returns are more unpredictable and exhibit less mean reversion. As such, when using them as a measure, we do not need to account for lagged values in of performance in our models. Stock returns are also an attractive performance measure because previous studies have shown that they are more predictive of executive turnover than accounting-based measures (Fee et al., 2015; Jenter and Kanaan, 2015). To adjust for the fact that stock return data has substantial outliers, I apply a log transformation.

### 2.3.4 Summary Statistics

In this section, I explore the differences and similarities between the personal and firm attributes of executives by gender. Perhaps unsurprisingly, male and female executives differ in their average age, length of tenure as an executive, and rank within their firm as determined by relative compensation. As seen in table 2.1, male executives are older, have longer tenure as a top executive, and are higher ranked than female executives in the sample. These results are largely consistent with findings in previous studies – e.g., Bertrand and Mullainathan (2001). Interestingly, while earlier studies found that female executives were more likely to be employed at smaller firms (Bertrand and Mullainathan, 2001), I find that female executives work at slightly larger firms.<sup>9</sup>

Importantly, however, there are not striking differences by gender in firm performance. While average annual returns seen by female executives is slightly lower than that seen by male executives, this difference is not statistically significant, suggesting that there are not substantial differences in the success of firms within the sample that employ male and female executives. Additionally, male and female executives do not experience different exposure rates to industry downturns.

Turning to the different measures of departure used in this study, there is a gender gap in executive departure as expected, given existing literature. The departure rate for female executives is approximately 3 percentage points higher than that for male executives. While an average departure rate of around 13% may seem high, this is quite close

---

data on departures in Execucomp is non-random.

<sup>9</sup>In addition to controlling for the number of employees to address the gender differences I observe in firm size, in my robustness checks I run a regression restricting the sample only to those firms that report between 4 to 6 executives in Execucomp (the most commonly observed range). The results largely remain unchanged.

Table 2.1: Summary Statistics

	Mean Female	Mean Male	Diff.
Age	54.17	57.36	3.20***
Tenure as Top Executive (Years)	6.29	7.13	0.84***
Rank (1 is highest)	3.80	3.19	-0.61***
Depart (%)	0.15	0.13	-0.02***
Involuntary (%)	0.15	0.12	-0.02***
Involuntary2 (%)	0.14	0.11	-0.03***
Annual Stock Return (%)	0.17	0.18	0.01
Number of Employees (Thousands)	19.69	18.46	-1.22*
Industry Downturns (%)	0.27	0.26	-0.00
N (Executive-Firm-Year)	6848	78152	
Unique Executives	1717	16787	

Notes: Tenure is measured as the number of years the executive is listed as a top executive at the firm. Rank is the executives relative rank within the firm as determined by total compensation. An executive is listed as having departed in a given year if they do not return to the firm for two consecutive years. An executive is listed as Involuntary in a given year if they satisfy the criteria for depart and do not reappear at any Execucomp firm within two years. An executive is listed as Involuntary2 in a given year if they satisfy the criteria for Involuntary and they are not marked as retired or deceased. \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

to other studies that look at CEO or executive departure. For instance, [Jenter and Kanaan \(2015\)](#) finds an annual CEO departure rate of 10.25% in their 1993-2009 sample; [Guay et al. \(2014\)](#) finds an annual CEO departure rate of 12% in their 1992 to 2008 sample; and [Fee et al. \(2015\)](#) finds an 11% annual CEO departure rate in their 1991-2007 sample. Keeping in mind that non-CEO executive departure rates tend to be somewhat higher than CEO departure rates, an average departure rate of 13% for top executives seems consistent with earlier studies.<sup>10</sup>

The difference in DEPART and INVOLUNTARY gives the unconditional reappearance rate for executives in the sample. The fact that this value is 1% suggests that only a small portion of the executives in the sample that depart reappear again at another Execucomp firm.<sup>11</sup> In fact, conditional on having departed, only around 7% of executives secure a new position with an Execucomp firm within two years of departure. While a conditional reappearance rate of 7% (or an unconditional reappearance rate of 1%) seems quite low, this is consistent with findings in existing studies. [Gayle et al. \(2012\)](#) finds an unconditional reappearance rate of 2% among top executives, while [Fee et al. \(2015\)](#) finds the CEO reappearance rate as a top executive conditional on having departed is 5.46%.

## 2.4 METHODOLOGY

Before discussing the details of the gender gap in executive departure rates, we must first define what constitutes a top executive position. Top executives are the highest ranked executives in a company and belong to the top management team (TMT) within the organization. As defined in ([Castanias and Helfat, 1991](#)), and in ([Hambrick and Mason, 1984](#)), a TMT consists of the CEO, and the senior executives that report directly to the CEO. These positions are typically at or above the level of vice president, including titles such as CFO, COO, etc. Unlike low-ranked employees and even middle management that can be downsized and their work reassigned to other employees, top executive positions are not disposable or interchangeable in the same sense. As such, dismissal decisions of TMT executives are unlikely to be due to cost-cutting by the board and are instead based on observed performance of top executives, as measured via firm performance.

Therefore, one method to identify whether the observed gender differences in executive departure are coming solely from differences in ability, is to strip away the ability component of firm performance and see if the exogenous portion of firm performance is still correlated with the departure of executives. To do this, I rely on the empirical framework created by the relative performance evaluation literature — i.e., [Holmström \(1982\)](#), [Jensen and Murphy \(1990\)](#), [Bertrand and Mullainathan \(2001\)](#), [Jenter and Kanaan \(2006\)](#).<sup>12</sup>

---

<sup>10</sup>This study's departure rates are not comparable to [Becker-Blease et al. \(2010\)](#) due to the fact that they did not account for the missing departure data in Execucomp and thus only looked at Execucomp reported departures.

<sup>11</sup>There are no significant gender differences in the likelihood of reappearing at another Execucomp firm conditional on having departed. 7% of departing male executives reappear and 6.2% of departing female executives reappear. These differences are not significant according to a two-sided t-test ( $p=0.29$ ).

<sup>12</sup>Under the relative performance evaluation (RPE) theory, optimal incentive schemes should depend only on the ability and/or actions of executives; any variation in industry level performance should be filtered out, and thus should not affect incentives. Papers focused on testing RPE

Suppose that the performance for firm  $f$ ,  $y_f$ , is a function of the firm's ability (and thus the top management team's ability)  $a_f$ , a firm-specific noise parameter  $u_f$ , and an industry wide noise parameter  $o$ , represented as follows.

$$y_f = a_f + o + u_f$$

If ability is the only factor leading to gender differences in departure, variation in firm performance that occurs at the industry level (i.e., variations in  $o$ ) are exogenous from ability for any given executive within the firm and thus should not lead to different performance evaluations for male versus female executives. In other words, industry performance (with the own-firm removed) should be orthogonal to the interaction between gender and involuntary executive departure within a firm. Note that this method relies on the assumption that the firm's actions do not affect the performance of industry peers; [Jenter and Kanaan \(2015\)](#) uses the same data set that I do and finds that this assumption holds.<sup>13</sup>

Given that we see a higher departure rate among female executives, I am especially interested in what happens to the gender difference in departures following contractions firm performance. Looking at the relationship between departures and a continuous measure of exogenous performance interacted with gender will show whether there is a stronger relationship between exogenous performance and departures for females versus males. This measure will not, however, directly show whether there are higher departures for females relative to males as a result of negative exogenous changes in performance. As such, much of the analysis in this study focuses on the relationship between departures and a binary measure of exogenous performance. By allowing the binary measure of performance to equal one only when the annual return is negative, it is possible to identify whether we specifically observe an increase in the gender difference in departure rates following industry contractions.

I focus on this measure because the change from an increase in annual stock value to a decrease relative to the prior year is a rather natural and salient cut-off. A large literature within psychology and behavioral economics on loss aversion demonstrates that the disutility individuals experience from a loss is typically larger than the utility experience from an equally sized gain.<sup>14</sup> As such, boards and investors may be more responsive to losses in firm value (as evidenced by a negative annual stock return) than gains. For robustness, I later consider an alternative definition of a negative shock to be instances where predicted firm performance is below the average industry mean performance across all years in the sample. This alternative definition can be motivated by a reference-dependence utility model in which the reference point is based on expectations about performance ([Koszegi and Rabin, 2006](#)). I show that the economic and statistical significance persists under this alternate specification.

---

construct instruments for firm performance that are exogenous from ability and look at how changes in instrumented firm performance affect various measures of executive incentives (e.g., bonuses, dismissals, etc.).

<sup>13</sup>The authors argue that were a firm's actions to affect the performance of industry peers, then the returns to industry performance for that firm should differ from returns for a firm that does not affect industry peer performance. Small firms within an industry are least likely to have an effect on the performance of industry peers whereas large firms are most likely to affect industry peer performance. [Jenter and Kanaan \(2015\)](#) shows that there is no difference to the returns for large versus small firms within an industry. [Bertrand and Mullainathan \(2001\)](#) also provides evidence in favor of this assumption.

<sup>14</sup>Loss aversion was first introduced in seminal work by [Kahneman and Tversky\(1979\)](#). For a recent review of the loss aversion literature see [DellaVigna\(2009\)](#) or [Marzilli Ericson and Fuster\(2014\)](#)

More formally, I estimate departure probability for a given executive at a firm within a year as a function of the executive's gender, an indicator for whether firm performance is negative, the interaction between gender and an indicator for negative performance, and additional controls. Letting  $i$  represent executives,  $f$  represent firms,  $d$  industries, and  $t$  years, the estimating equation can be defined as follows.

$$depart_{i,f,t} = \beta_0 + \beta_1 NEG_{f,t} + \beta_2 F_i + \beta_3 NEG_{f,t} \times F_i + \theta \mathbf{X}_{i,f,t} + \delta_t + \varphi_d + \epsilon_{i,f,t}$$

$F_i$  is a dummy variable that takes the value of 1 if the executive is female,  $NEG_{f,t}$  is a dummy variable that equals 1 when the firm's annual return is negative, and  $\mathbf{X}_{i,f,t}$  is a vector of controls for executive  $i$  at time  $t$ .<sup>15</sup> Standard errors are clustered at the industry level to account for the fact that performance is highly correlated across firms within a given industry. If gender differences in departure stem solely from differences in ability, we should find  $\beta_3 = 0$ . If, however,  $\beta_3 \neq 0$ , then the gender difference in executive departures is at least partially coming from a factor other than gender differences in ability.

Because firm performance is a function of both executive ability and exogenous factors, I instrument for negative firm performance using an indicator for when own-firm-removed industry performance is negative. As is standard when dealing with an interaction between an exogenous variable (Gender) and an endogenous variable (negative firm performance), I use the interaction between the exogenous variable and the instrument for the endogenous variable as an instrument for the interaction term. That is, I use the interaction between gender and an indicator for when own-firm-removed industry performance is negative to instrument for the interaction between gender and negative firm performance. More formally, for a given firm  $f$  in year  $t$  with  $n$  firms in the industry let own-firm-removed industry performance for firm  $f$ ,  $performance_{f,t}^{Industry}$  be defined as follows:

$$performance_{f,t}^{Industry} = \frac{1}{n-1} \sum_{j \in Industry_{f,t} \text{ s.t. } j \neq f} performance_{j,t}$$

From this definition,  $1\{performance_{f,t}^{Industry} < 0\}$  is used as an instrument for  $NEG_{f,t}$  and its interaction with executive gender is used as an instrument for  $NEG_{f,t} \times F_i$ . Although the endogenous variables used are both binary treatment variables, I employ a garden variety 2SLS methodology. This is the recommended course of action suggested by Angrist (2009).

---

<sup>15</sup>Controls include the number of employees within the firm, the executive's rank in the firm as determined using the executive's annual compensation relative to the other top executives within the firm, tenure as a top executive within the firm, and age. I include time fixed effects and industry fixed effects in all specifications except those without controls.



## 2.5 RESULTS

### 2.5.1 Main Result

Before discussing the main regression output, I first look at Figure 2.1, which plots the relationship between predicted firm performance (as computed using the first stage regression) and my preferred measure of executive resignations that excludes executives who move to new firms (i.e. INVOLUNTARY). This figure shows that, while departure rates increase for both male and female executives when firm performance is predicted to be negative, males only experience a 1.3 percentage point increase in departure whereas females experience a 3.8 percentage point increase in departure. Although this figure is suggestive of female executives having a higher likelihood of departure as a result of industry contractions, it does not control for any differences in executive characteristics or firm characteristics. Thus, to better understand the effect of industry contractions on gender differences in executive departure, we need to turn to our second stage regression results.

Figure 2.1: Departure by Gender and Exogenous Performance

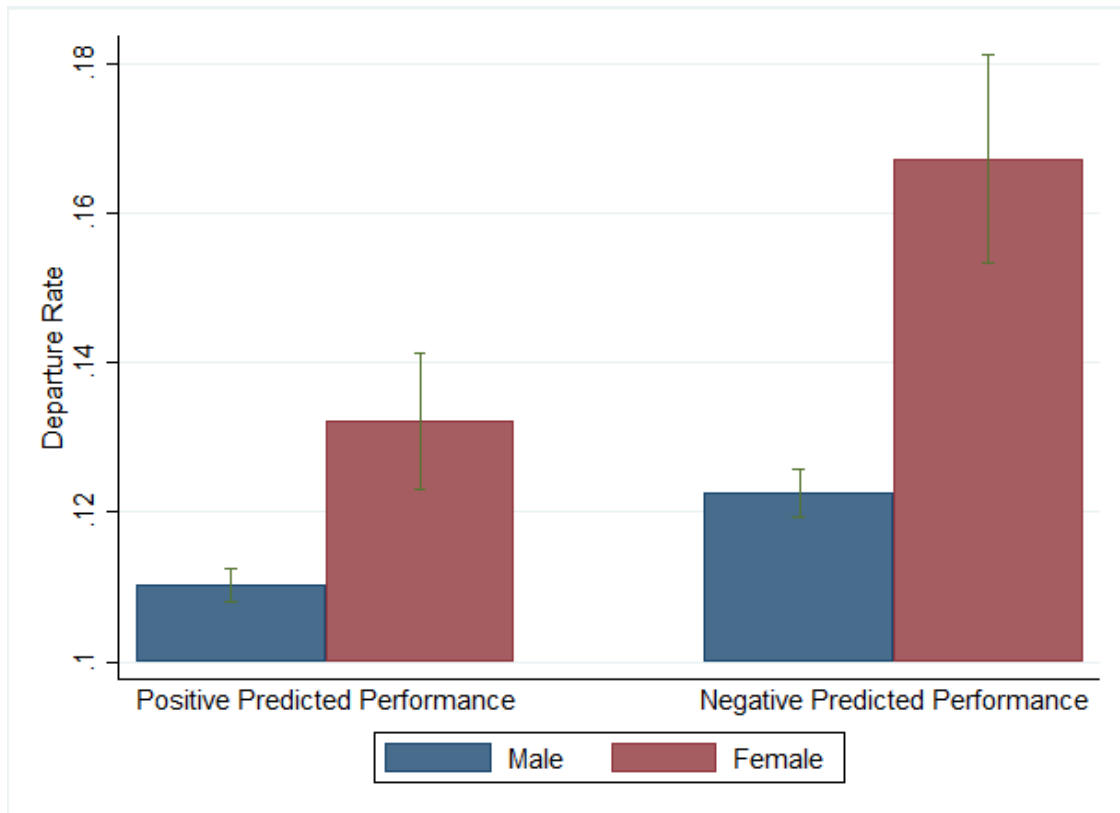


Table 2.2 presents the results from a 2SLS regression in which I instrument for negative firm performance and its interaction with gender. The first three columns report results from the second stage regression using my preferred

measure of departure as the dependent variable. Across these three columns, regardless of whether we have neither controls nor fixed effects, only fixed effects, or both controls and fixed effects, the estimated gender gap in departure following downturns (measured by Negative X Female) remains quite stable.<sup>16</sup> Furthermore, the remaining two columns report results using my alternate two measures of departure. For all three measures of departure, female executives are predicted to have a higher overall resignation rate than male executives following industry downturns. We can also see that overall departure rates for female executives are positive and significant. This matches what is shown in Figure 2.1.

Focusing now on my preferred measure of departure and specification in column 3, the estimated coefficients suggest that overall resignation rates for female executives are similar to that for male executives. When firm performance is negative however, the resignation probability of female executives increases by around 5 percentage points more than male executives. Recall from table 2.1 that, as measured by INVOLUNTARY, the average departure rate for female executives is 15%. Thus, the estimated increase in departure rates of female executives during contractions is both statistically significant and economically significant.

Recall, INVOLUNTARY does not code as departed those executives that leave their job and move to another Execucomp firm; INVOLUNTARY2 includes the same restrictions as INVOLUNTARY, but does not code as departed those executives that Execucomp flags as being retired or deceased. Thus, INVOLUNTARY2 is a tighter definition of departures. In contrast, DEPART codes all executives that leave their job as departed. Thus, DEPART is a looser measure of departure than INVOLUNTARY. Looking across column 4 and column 5 of Table 2.2, the estimated treatment effect decreases in magnitude under the looser measure of departure and increases in magnitude under the tighter measures of departure. This suggests that the increase in female departure rates does not appear to be being driven by voluntary job changes. I am not, however, unable to rule out other voluntary explanations for the increase in departures (such as leaving the market to become a stay-at-home mom). I explore these alternative explanations in more detail in section 2.6.

Regardless of whether the increase in the gender gap in departures following exogenous decreases in firm performance is being driven by voluntary or involuntary departures, the fact that we see the gender gap in departures increasing following an exogenous change in firm performance allows us to conclude that the gender gap in departures is not driven solely by gender differences in ability. In section 2.6, I will discuss potential mechanisms that may be driving this effect. For the remaining discussion, unless stated otherwise, I will focus on my preferred measure of departure, INVOLUNTARY.

---

<sup>16</sup>Interestingly, the same is not true for male executives. While the male departure rate is predicted to increase following downturns in column 1, this effect is no longer significant after adding year and industry fixed effects. One potential interpretation for this is that male executives experience increased rates of departures following overall recessions; however, they do not experience increased departure rates from industry-specific downturns.

Table 2.2: Departure Probability:  
Results of Second-Stage Regressions

	(1)	(2)	(3)	(4)	(5)
	Involuntary	Involuntary	Involuntary	Depart	Involuntary2
Negative	0.023*** (0.005)	0.017 (0.020)	0.018 (0.018)	0.017 (0.020)	0.021 (0.018)
Negative X Female	0.050** (0.022)	0.051** (0.021)	0.050** (0.020)	0.037* (0.020)	0.056*** (0.020)
Female	0.012 (0.008)	0.007 (0.007)	0.003 (0.007)	0.008 (0.008)	0.001 (0.007)
Year FE	No	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes
N	84998	84998	84998	84998	84998

Notes: Performance is the predicted log annual stock return from the first stage regression using mean industry performance with the own-firm removed. Negative is a dummy variable that equals 1 when predicted performance is negative. Controls include executive age, rank, tenure, and number of employees. Robust standard errors (clustered at the industry level) are reported in parentheses. \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

### 2.5.2 Robustness Checks

I conduct a number of robustness checks in which I vary the specification, the sample restrictions, and check for omitted variable bias. Overall, the results are largely unchanged. In this subsection, I briefly discuss the robustness checks conducted and their findings.

The main specification in this study only allows for a discrete jump in departure rates following negative firm performance. One question is whether the results will change if an interaction between the continuous measure of performance and gender is included. Table 2.3 presents results from this specification and shows that the relationship is indeed stronger for female executives and in order for positive annual returns to counteract the fact that female executive departure rates are predicted to be higher than that for males, the exogenous portion of log annual return would need to exceed a value of 1.

One potential concern is that the estimated treatment effect is being driven solely by differences near the discontinuity. Ideally, the observed increase in the gender gap in departure should hold even when we move away from this discontinuity point. To address this concern, I estimate the second stage regression excluding observations in a narrow band around the discontinuity. Columns 2 through 4 of Table 2.4 present results that exclude all observations in which firm performance is within 0.025, 0.05, and 0.075 of 0. In doing this, the results are largely unchanged. A related concern is that, while zero is a natural and salient cut-off to use when defining negative performance, other cut-offs may also be worth considering. Column 5 presents results from a regression that redefines a negative shock to being instances in which performance is below historic mean returns. In other words, the cutoff is moved from being at zero to being at the overall mean historic return. While the treatment effect decreases some in magnitude, it remains both economically and statistically significant. This suggests that the results are not being driven only by large differences around the discontinuity nor are they being driven by the choice of cut-off.

Concerns may also arise regarding sample restrictions. Recall that in the primary specification, executives that have a start year as an executive that is before 1997 are excluded from the sample. Column 2 of table 2.5 shows that after relaxing this assumption, the effect remains economically and statistically significant. As an additional test, recall that there exist some gender differences in firm size and that the number of reported executives varies some by firm. One concern stemming from this is that there may be differences in the firing behavior of large versus small firms that are driving the treatment effect. To address this, column 3 of Table 2. 5 shows that that the primary treatment effect persists even when the sample is restricted to only include firms that report between 4 to 6 executives (the range into which most firms fall).

Finally, I check for selection on unobservables stemming from omitted variables. Execucomp does not provide information on the background of executives. To address this, I first compare the primary regression to one without controls. Adding controls does not have a large effect on the estimated treatment effect. To more rigorously test for the robustness of the results, I use the method detailed in [Oster \(2016\)](#) to test how much variation in departures would need to be explained by unobservables in order to reduce the treatment effect to zero. The standard rule of thumb is to

Table 2.3: Departure Probability with  
Continuous Performance Measures:  
Results of Second-Stage Regression

	(1)
Performance	-0.009 (0.012)
Performance X Female	-0.029* (0.017)
Female	0.027*** (0.005)
Year FE	Yes
Industry FE	Yes
Controls	Yes
N	84998

Notes: Performance is the predicted log annual stock return from the first stage regression using mean industry performance with the own-firm removed. Controls include executive age, rank, tenure, and number of employees. Robust standard errors (clustered at the industry level) are reported in parentheses. \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table 2.4: Sensitivity of Departure Probability to Measure of Industry Shocks:  
Results of Second Stage Regressions

	(1)	(2)	(3)	(4)	(5)
	Baseline	Exclude +/- .025	Exclude +/- .05	Exclude +/- .075	Alternate Cut-Off
Negative	0.018 (0.018)	0.023 (0.018)	0.025 (0.018)	0.028 (0.018)	
Negative X Female	0.050** (0.020)	0.042* (0.022)	0.043** (0.022)	0.040* (0.021)	
Female	0.003 (0.007)	0.009 (0.008)	0.008 (0.008)	0.010 (0.008)	0.007 (0.008)
Below Average					0.007 (0.014)
Below Average X Female					0.032* (0.017)
Year FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
N	84998	80343	75789	71452	84998

Notes: Performance is the predicted log annual stock return from the first stage regression using mean industry performance with the own-firm removed. Columns 2 through 4 exclude observations with predicted performance in the specified interval around 0. Column 5 reports results from a regression in which the definition of a negative shock is changed to be instances in which predicted performance is below the overall industry mean performance with the own-firm removed. Negative is a dummy variable that equals 1 when predicted performance is negative. Below Average is a dummy variable that equals 1 when predicted performance is below the overall industry mean performance with the own-firm removed. Controls include executive age, rank, tenure, and number of employees. Robust standard errors (clustered at the industry level) are reported in parentheses. \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table 2.5: Additional Robustness Tests: Results of  
Second-Stage Regressions

	(1)	(2)	(3)
	Baseline	Include Pre-1997	Firms with 4-6 Executives
Negative	0.018 (0.018)	0.021* (0.011)	0.000 (0.017)
Negative X Female	0.050** (0.020)	0.042** (0.018)	0.048** (0.020)
Female	0.003 (0.007)	0.004 (0.007)	0.001 (0.007)
Year FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
N	84998	118982	72653

Notes: Performance is the predicted log annual stock return from the first stage regression. Negative is a dummy variable that equals 1 when predicted performance is negative. Controls include executive age, rank, tenure, and number of employees. Robust standard errors (clustered at the industry level) are reported in parentheses. \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

carry out this estimation with the assumption that the maximum possible R-squared (if all unobservables are included) is 1.3 times the R-squared observed in the regression with controls. Using this rule of thumb, estimates suggest that unobservables would need to explain over 160 times more variation in departures than observables in order to render the treatment effect null. As a more conservative test, I repeat the Oster test but instead assume the maximum possible R-squared is 1. Using this maximum R-squared, unobservables would need to explain 4.74 times more variation in departures than observables. Given that the standard cutoff for this test is 1, I can confidently reject the hypothesis that the treatment effect is being driven by omitted variable bias.

## **2.6 MECHANISMS**

So far, I have demonstrated that the gender gap in executive departure increases following industry downturns, and that this increase is driven solely by a change in the departure rate of female executives. Given that industry performance is exogenous from executive ability, the data suggest that this increase is not due to gender differences in ability.

This section discusses alternative explanations for the observed treatment effect. Given that the measure of departures used cannot rule out voluntary departure from the market of Execucomp firms, I first test whether fertility preferences or early retirement can explain the increases in gender gap in departures following industry contractions. Neither fertility nor early retirement appear to explain this increase. Next, I consider whether differences in hiring conditions might account for the increased rate of female departures and demonstrate that this also does not appear to be driving the observed effect. The subsequent section then discusses differences in attribution bias or blame as a plausible remaining explanation.

### **2.6.1 Fertility**

In order to consider the total effect industry contractions may have on fertility behavior of female executives, we must separately consider the substitution and income effect. An industry-wide contraction will likely affect existing female executives' pay through reduced bonuses, as well as reduced stock values. This reduction in pay will reduce the relative cost of leaving the industry for child-rearing purposes and thus, via the substitution effect, executives will shift away from work and towards child-rearing. However, the decrease in pay will also reduce the overall income of executives and thus, via the income effect, will cause executives to work relatively more. Depending on whether the income or substitution effect dominates, we could see industry contractions increase, decrease, or not change the rate of departure for fertility reasons of female executives.

Looking at the relationship between the overall labor market and economic fluctuations, it seems unlikely that we would see an increase in fertility decisions by executives during industry contractions. Overwhelmingly, studies focused on the relationship between business cycles and fertility find that fertility rates decrease during recessions



and are thus pro-cyclical (refer to [Jones and Schoonbroodt \(2016\)](#), [Schneider \(2015\)](#) for studies using recent data, and [Sobotka et al. \(2011\)](#) for an extensive survey of earlier studies). Furthermore, [Schaller \(2016\)](#) finds that the substitution effect is weakest among high-skilled women, suggesting that fertility rates should decline most among these women during economic contractions. Thus, in order for fertility to explain the increase in female executive departure rates during industry contractions, female executives would need to differ substantially from not only the average female worker, but also from the average high-skilled female worker.

The data provides further evidence against fertility being the primary mechanism behind the observed gender difference in departure rates. If female executives depart more for fertility reasons during economic contractions, then there should be a significant decrease in the magnitude and significance of the treatment effect when the sample is restricted to exclude younger executives that are more likely to be making fertility decisions. Table 2.6 presents results when the sample is restricted to executives over the age of 40 (column 2), over the age of 45 (column 3), and over the average menopausal age of 51 (column 4). Contrary to what one would expect if economic contractions are leading women to become stay-at-home moms, the effect of downturns on female executive departure remains both economically and statistically significant.

As additional tests for fertility as a mechanism, I explore the overall trends for executives by gender at the market level and look at what happens around the two recessions observed in the sample (2001 and 2007-2009). If fertility is driving the increase in departures during downturns, then the share of entrants that are women should decrease during recessions, as women move away from the labor market for fertility. Figure 2.2 shows that this is not occurring and that, if anything, the share of female entrants increased during the recessions. One may also expect to see the age of female entrants increase during recessions as younger women choose not to enter the market for fertility reasons; however, as Figure 2.3 demonstrates, there is not a consistent relationship between the change in the age of female entrants across the two observed recessions. Finally, the average age of departing female executives decline during recessions if women in their fertile years leave the labor market. Figure 2.4 shows that there is no consistent relationship between the age of departing or non-departing female executives and recessions.

Overall, the data do not provide any evidence that is consistent with the fertility mechanism driving the treatment effect. Given the lack of evidence in favor of the fertility mechanism, in combination with the fact that existing literature suggests fertility is pro-cyclical, fertility does not appear to be the main factor driving increased female departure during industry downturns.

### **2.6.2 Early Retirement**

Another possibility is that the observed increase in female departure rates following downturns is coming from early retirement by female executives. As with fertility, we need to consider both the income and substitution effects of industry downturns. While industry downturns will reduce the relative price of early retirement (as discussed in the fertility section), they also reduce the overall income of executives. In order to be the mechanism driving increased

Table 2.6: Subsample Analysis using Older Executives:  
Results of Second Stage Regressions

	(1) Baseline	(2) Age > 40	(3) Age > 45	(4) Age > 51
Negative	0.018 (0.018)	0.019 (0.019)	0.015 (0.020)	0.022 (0.025)
Negative X Female	0.050** (0.020)	0.049** (0.021)	0.058*** (0.021)	0.088*** (0.031)
Female	0.003 (0.007)	0.004 (0.008)	0.004 (0.008)	-0.002 (0.012)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
N	84998	82643	75415	56500

Notes: Performance is the predicted log annual stock return from the first stage regression using mean industry performance with the own-firm removed. Negative is a dummy variable that equals 1 when predicted performance is negative. Controls include executive age, rank, tenure, and number of employees. Robust standard errors (clustered at the industry level) are reported in parentheses. \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Figure 2.2: Percentage of Female Entrants by Year

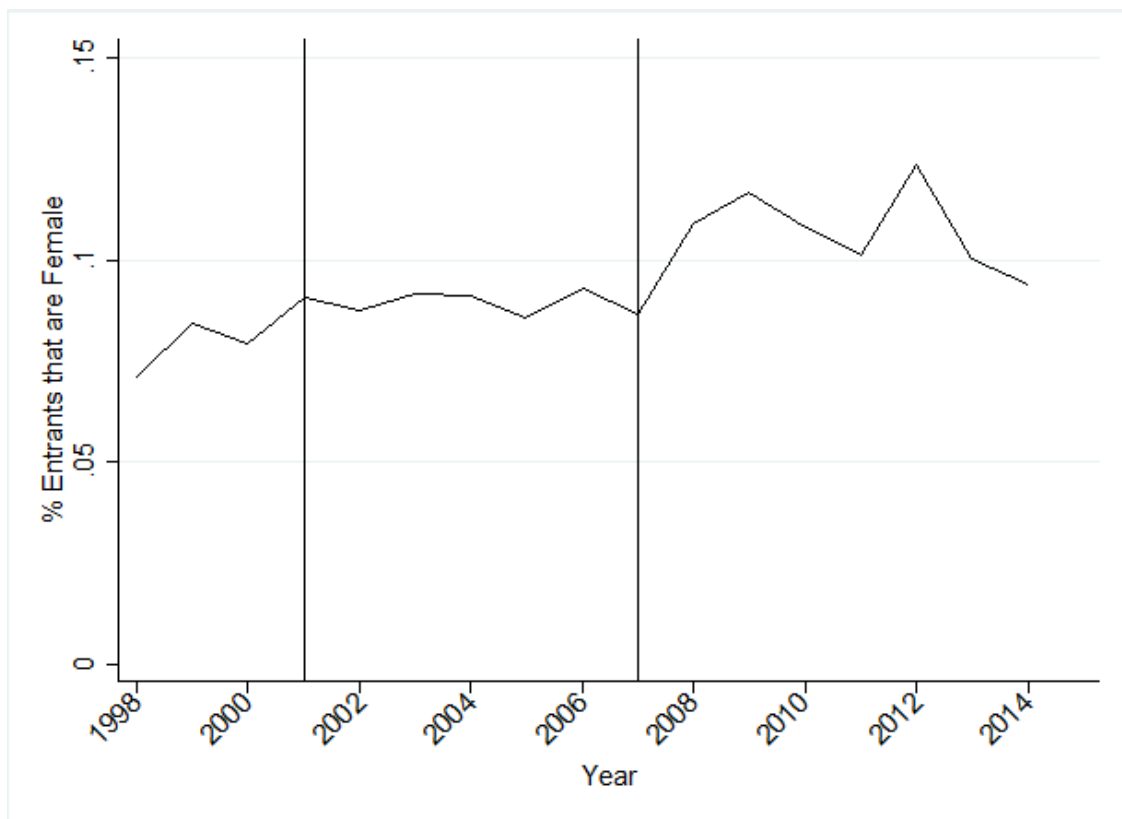
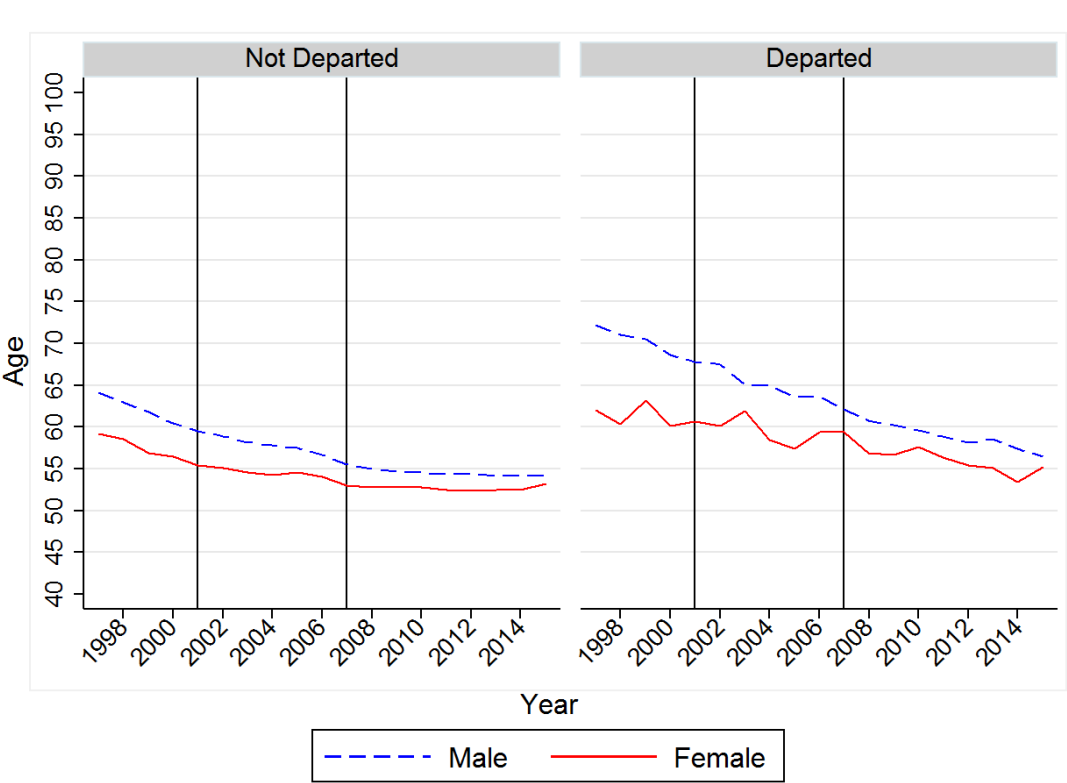


Figure 2.3: Average Age of Entrants by Year and Gender



Figure 2.4: Average Age of Executives by Year, Gender, and Departure Status



female departures for during downturns, it would need to be the case that the substitution effect dominates the income effect for female executives at retirement age, but not for male executives at retirement age. Furthermore, pension payments for CEOs and top executives are based on pay around the point of retirement. Liebersohn (2016) shows that executives take this into account when timing their retirements and, as a result, retirement among executives is lower during economic contractions relative to expansions.

Despite the fact that early retirement does not appear to be a likely explanation for the observed treatment effect given the results in Liebersohn (2016), I investigate it in more detail via subsample analysis and observation of trends for executives by gender at the market level. As a first step, I consider what happens to the gender gap in departures following downturns when I restrict the sample to executives below retirement ages. If early retirement is the mechanism driving increased female departures following downturns, the treatment effect should be reduced in magnitude and significance. As Table 2.7 shows, this does not occur. Restricting the sample to executives below the classic retirement age of 65 or the younger age of 60, we see that the treatment effect persists, suggesting early retirement is not the mechanism driving the increased gender gap in executive departures.

To provide additional evidence against the early retirement mechanism, recall that we did not see any systematic relationship between executive age and recessions (Figure 3 and Figure 4). Thus, there is no evidence that early retirement is driving the increase in female departure rates during downturns.

### 2.6.3 Glass Cliff

Another possible explanation for the gender differences in executive departure during downturns is the glass cliff. This term was first introduced in Ryan and Haslam (2005), in response to an article in *The Times*, which suggested that female leaders negatively affect firm performance. Ryan and Haslam refute this claim and argue that women are more likely to be appointed into leadership positions during times of market contractions when firm performance relative to other firms is poor. Thus, female leaders have a higher likelihood of being appointed to more tenuous positions. Ryan and Haslam refer to this phenomenon as the glass cliff.<sup>17</sup>

While several studies have found evidence of the glass cliff phenomenon (e.g., Kulich et al. (2015); Mulcahy and Linehan (2014); Haslam and Ryan (2008)), there is also evidence suggesting that it may not persist in all settings and may be dependent on the sampling population and definition of “crisis” within a firm (Adams et al., 2009; Cook and Glass, 2014; Hennessey et al., 2014). Furthermore, much of the research on the glass cliff phenomenon focuses on female board members and CEOs. Thus, while it is possible the glass cliff is contributing to the gender differences in departure, it is also possible that the glass cliff is not a strong presence within the population of all top executives at publicly traded US firms.

To explore whether the glass cliff is driving the increase in departure probability of female executives during

---

<sup>17</sup>Reasons for the glass cliff phenomenon include a desire to signal change to investors (Kulich et al., 2015), selection bias (Haslam and Ryan, 2008), and implicit stereotypes that female executives are better in times of crisis – also known as, “think crisis – think female” (Ryan et al., 2011).

Table 2.7: Subsample Analysis using Younger Executives:  
Results of Second Stage Regressions

	(1)	(2)	(3)
	Baseline	Age < 65	Age < 60
Negative	0.018 (0.018)	0.017 (0.017)	0.015 (0.019)
Negative X Female	0.050** (0.020)	0.049** (0.022)	0.054*** (0.019)
Female	0.003 (0.007)	0.000 (0.007)	-0.003 (0.006)
Year FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
N	84998	73608	60443

Notes: Performance is the predicted log annual stock return from the first stage regression using mean industry performance with the own-firm removed. Negative is a dummy variable that equals 1 when predicted performance is negative. Controls include executive age, rank, tenure, and number of employees. Robust standard errors (clustered at the industry level) are reported in parentheses. \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

downturns, I restrict my attention to the subsample of executives that were hired when firms were doing better than their industry peers (i.e., when the residual from the first stage regression is positive), and also to the subsample of executives that were hired when firms were doing better than their industry peers for two consecutive years. These subsamples represent the set of executives that are least likely to be glass cliff hires for females, and thus we should no longer observe the treatment effect if the glass cliff is driving the results. Table 2.8 shows that this is not the case. Even among the set of executives hired in firms that are doing relatively well, female executives' departure probability increases during downturns.

As a secondary test of whether the glass cliff hypothesis is driving the treatment effect, I consider a subsample of executives with longer tenure. Given that glass cliff executives are hired into more tenuous positions, one may expect that their expected tenure is much shorter than that of non-glass cliff hires. As such, the effect of industry downturns on female departures should be weaker among executives with longer tenure. This subsample analysis is slightly trickier to interpret, because executives with longer tenure are already, through selection, likely to be of better quality, and thus less sensitive to shocks. Regardless, focusing on Table 2.9, it only after the sample nearly in half – by restricting our attention to executives with more than 3 years of tenure – that the treatment effect declines in magnitude and significance. One would imagine that, if the glass cliff was the driving mechanism behind the increased departure rate of female executives during downturns, we would see the estimated treatment effect with a much earlier tenure restriction. Thus, the results from Table 2.9 also provide evidence against the glass cliff hypothesis.

Taking the results from Table 2.8 and the results from Table 2.9, the glass cliff hypothesis does not appear to be the mechanism behind the treatment effect. Thus, while female executives may be more likely to be hired into already tenuous positions, even those female executives hired into less tenuous positions experience an increase in departure probability following industry downturns. This suggests that regardless of whether female executives face a glass cliff or a solid ledge ([Hennessey et al., 2014](#)), both experience a precipitous drop when the ground starts to shake.

#### **2.6.4 External Hires**

There exists some evidence that, conditional on being appointed to a CEO position, females are more likely to come from an external appointment than males. In a sample of CEOs from 2004 to 2015, 77% of male appointments were internal, whereas only 68% of female appointments were internal ([Study, 2016](#)). Expanding to the set of all top executives, however, the picture is less clear. [Quintana-García and Elvira \(2017\)](#) finds that, in a sample of technology firms, female executives are actually less likely to be externally hired than male executives. In contrast, [Fernandez-Mateo and Fernandez \(2016\)](#) finds that, in a proprietary sample of firms that rely partially on an external hiring agent, female executives are more likely to be externally hired. One could expect that this may lead to those executives having weaker internal networks and thus being more exposed to exogenous declines in firm performance.

There is mixed evidence on whether being externally hired is detrimental to executive success. While on the one hand, externally hired executives tend to experience higher starting compensation than internal hires ([Bidwell,](#)



Table 2.8: Subsample Analysis using Executives Hired when Firm is Performing Well Relative to Peers: Results of Second-Stage Regressions

	(1)	(2)	(3)
	Baseline	Industry-Adjusted Return Positive when Hired	Industry-Adjusted Return Positive when Hired and Prior Year
Negative	0.018 (0.018)	0.040** (0.017)	0.050* (0.028)
Negative X Female	0.050** (0.020)	0.085*** (0.026)	0.079** (0.031)
Female	0.003 (0.007)	-0.010 (0.007)	-0.014* (0.009)
Year FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
N	84998	51376	35111

Notes: Performance is the predicted log annual stock return from the first stage regression using mean industry performance with the own-firm removed. Negative is a dummy variable that equals 1 when predicted performance is negative. Controls include executive age, rank, tenure, and number of employees. Robust standard errors (clustered at the industry level) are reported in parentheses. \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table 2.9: Subsample Analysis using Executives with More Tenure:  
Results of Second-Stage Regressions

	(1)	(2)	(3)	(4)
	Baseline	Tenure > 1 year	Tenure > 2 years	Tenure > 3 years
Negative	0.018 (0.018)	0.027 (0.017)	0.018 (0.020)	0.020 (0.022)
Negative X Female	0.050** (0.020)	0.051** (0.022)	0.069*** (0.022)	0.041* (0.024)
Female	0.003 (0.007)	-0.006 (0.008)	-0.013 (0.009)	-0.004 (0.009)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
N	84998	71472	58579	45310

Notes: Performance is the predicted log annual stock return from the first stage regression using mean industry performance with the own-firm removed. Negative is a dummy variable that equals 1 when predicted performance is negative. Controls include executive age, rank, tenure, and number of employees. Robust standard errors (clustered at the industry level) are reported in parentheses. \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

2011), this benefit is generally limited to males (Brett and Stroh, 1997; Dreher et al., 2011). In fact, Quintana-García and Elvira (2017) argues that externally hired female executives may fare worse with respect to compensation than internally hired female executives due to more discretion in pay setting for external hires. It is unclear, however, to what extent the results in Quintana-García and Elvira (2017) will generalize to executive departure, given that the disadvantage they find for female external hires stems from gender differences in pay-setting behavior.

Focusing more narrowly on departures, neither Dikolli et al. (2014) nor Yonker (2016) find a significant relationship between being externally hired and departure probability. In contrast, Guay et al. (2014) finds that externally hired CEOs experience higher overall involuntary departure; importantly, however, this study also finds that externally hired CEOs are less likely to involuntarily depart following exogenous industry shocks. Thus, if anything, being externally hired would suggest a decrease in the departure probability following industry downturns.

Considering the overall literature on gender differences in external hire rates and the literature on the impact of being externally hired on labor market outcomes, it seems unlikely that differences in external hiring could be driving the gender differences in departure rates. Specifically, it is unclear whether, when considering all top executives, there exists a gender difference in external hire probabilities. It also appears that, even conditional on being an external hire, negative exogenous shocks to firm performance should at worst have no effect on the departure rates of external hires.

Unfortunately, identifying internal versus external hires in the sample is difficult. For many executives within the sample, there is missing information on when they first joined their company.<sup>18</sup> Importantly, however, looking at executives in the sample for whom I do have information on their join year, I do not observe an overall difference between male and female executives in the likelihood of being an external hire; 55.4% of males are externally hired whereas 53.7% of females are externally hired.<sup>19</sup> Thus, taking this in conjunction with the existing literature, this seems to be an unlikely mechanism for the observed treatment effect.

### 2.6.5 Female Start-Ups

One may also think a factor driving the observed treatment effect is a preponderance of female executives leaving their current positions to create their own startups during downturns. The argument would be that, due to the downturn, their salaries, pensions, and other perquisites are smaller, and the opportunity cost of leaving their current job is therefore reduced. However, in order for this explanation to be likely, the substitution effect of leaving one's job during a downturn to pursue a startup would have to outweigh the income effect of the downturn. Furthermore, (Ewing Marion Kauffman Foundation, 2017) finds no evidence of a higher number of female-created startups during recessions relative to the number of male-created startups. Therefore, it seems unlikely that the treatment effect is being driven by female executives leaving their positions to create startups during downturns.<sup>20</sup>

<sup>18</sup>Within the sample, this information is missing for 71.9% of the executives and is missing more often for female executives (76.4%) than male executives (71.5%).

<sup>19</sup>I observe a higher share of being externally hired for female CEOs that closely matches the Study (2016).

<sup>20</sup>One may also be tempted to think that the increased departure rates for female executives during downturns could be attributed to gender differences in the size of golden parachutes – the pension and severance payments top executives receive upon termination. This relies on the

## 2.7 MISPLACED BLAME

If the increase in departure rates among female executives following downturns is not due to differences in leadership ability, fertility, retirement, the glass cliff, or differences in external hire rates, what could explain this treatment effect? One remaining explanation that is consistent with the results presented within this study, and supported by related findings within the literature on gender differences in labor market outcomes, is misplaced blame.

It may be the case that boards incorrectly attribute poor firm performance to executive ability when the executive is female. In other words, it is possible boards may have a higher likelihood of committing fundamental attribution errors when performance is poor and the executive in question is female.<sup>21</sup> Misplaced blame could also manifest not as the result of attribution bias from the board, but rather via female executives being held to different performance standards (and thus punished more readily) relative to male executives. This could lead to female executives being blamed for outcomes they could not reasonably have control over or have foreseen. It may also be that the misplaced blame does not stem directly from the board, but could instead stem from CEO's being more likely to use female executives below them as scapegoats for poor performance. When considering the latter possibility, it is relevant to refer to the findings discussed in [Finkelstein et al. \(2009\)](#) that suggest performance does not provide a complete explanation for executive turnover. Indeed, [Finkelstein et al. \(2009\)](#) conclude powerful but underperforming CEOs will shift blame to their TMT executives, resulting in the replacement of those executives rather than the CEO. This behavior may also extend to gender whereby female executives may be blamed more often than males.

There are existing results that provide some evidence in support of women more frequently receiving misplaced blame. [Selody \(2010\)](#) and [Albanesi et al. \(2015\)](#) demonstrate that female executives' pay decreases more following negative changes in firm performance than does the pay for male executives, whereas male executive pay increases more than female pay following positive changes in firm performance. [Selody \(2010\)](#) further documents that male executives' bonuses increase more following exogenous increases in firm outcomes than do bonuses for female executives. Not only has research documented gender asymmetries in the extent to which executives are penalized or rewarded for their performance, there is also recent evidence documenting how women may be more harshly punished for blatant misconduct. Looking at financial advisors instead of executives, [Egan et al. \(2017\)](#) finds that women are more likely to be terminated for misconduct than men, despite the fact that women are less likely to engage in misconduct and are less likely to be repeat offenders. These results suggest that male and female employees receive differential treatment for successes and failures.

It is also possible that the differences in blame may not stem from the board or CEO, but may instead be due to gender differences in self-attribution bias. Experimental economics literature has not found gender differences in the

---

argument that the board would choose to eliminate an executive's position as a cost-cutting measure during downturns. However, as discussed previously, unlike other positions, TMT are unlikely to be made redundant. As such, it is unclear what benefit firing an existing executive simply to replace them would serve.

<sup>21</sup>Existing literature on attribution bias has demonstrated this phenomenon with respect to perceived leaders whereby individuals exhibit tendencies to over-attribute the role of the leader in both positive and negative exogenous outcomes ([Weber et al., 2001](#)). Attribution bias may also manifest more in bad times and more often among outsiders (such as among women in top corporate roles during downturns) ([Selody, 2010](#)).

direction of asymmetric updating within ego-relevant tasks – e.g., [Eil and Rao \(2011\)](#), [Ertac \(2011\)](#), [Möbius et al. \(2014\)](#). However, looking instead to the psychology literature, several studies show women are more likely to attribute failure to internal factors and success to external factors, while men behave in the opposite manner – e.g., [Deaux \(1979\)](#), [Deaux and Emswiller \(1974\)](#). This difference in self-attribution behavior may cause female executives to take the blame for failures that are outside of their control, and thus lead to increased departure.

Identifying blame as a mechanism for increased departure, and separating attribution error on behalf of the board from attribution error on behalf of executives, is not possible given the existing data set. Thus, while the results are consistent with blame as a possible mechanism, I cannot rule out other mechanisms. Future research may wish to use experimental methodology in order to address these questions.

## 2.8 CONCLUSION

The underrepresentation of women in top-management positions has received substantial attention within the economics literature. While most studies have focused on differences in entry into management positions, a smaller set of studies have instead looked at departure. These studies have largely shown that female executives exhibit higher departure rates.

Why is it that female executives exhibit higher departure rates than male executives? One can think of many possible answers to this question. Perhaps the simplest explanation would be gender differences in ability. In order to test for this, one must examine gender differences in departure resulting from exogenous changes in firm performance.

In this study, I use exogenous changes in firm performance to rule out the ability channel and show that gender differences in departure still persist. Specifically, using industry performance as an instrument for firm performance, I show that following industry-wide contractions, the overall departure rate for female executives increases by approximately 5 percentage points, while no change is observed in the departure rate for male executives. This increase in departure cannot be explained by differences in fertility, early retirement, the glass cliff, or differences in external hire rates. I argue that one remaining channel that is consistent with the observed increase in departure rates is misplaced blame (e.g., attribution bias, self-attribution bias, or unreasonable performance standards).

While the results are consistent with a blame hypothesis, other competing mechanisms may remain. As such, one avenue for future related research is using experimental methodology to test the blame hypothesis and separate out self-attributed blame from external blame.

**3.0 THE SLIDER TASK: AN EXAMPLE OF RESTRICTED INFERENCE ON INCENTIVE EFFECTS (CO-AUTHORS: FELIPE A. ARAUJO, ERIN CARBONE, LYNN CONELL-PRICE, MARLI W. DUNIETZ, ANIA JAROSZEWICZ, DIEGO LAMÉ, LISE VESTERLUND, STEPHANIE W. WANG, AND ALISTAIR J. WILSON)**

**3.1 INTRODUCTION**

Early economic experiments examining labor effort in the lab relied on the stated-effort design (for example: [Bull et al., 1987](#); [Schotter and Weigelt, 1992](#); [Nalbantian and Schotter, 1997](#); [Fehr et al., 1993](#))<sup>1</sup>. Participants in the role of workers were given an endowment and asked to “purchase” a level of effort, which in turn benefited other participants in the role of principals. While stated-effort designs provided well-structured controls for participants’ costs of effort, the designs were seen as being abstract, overly distant from the types of labor effort the experiments were intended to capture. Scholars subsequently began to use real-effort designs, where participants are instead paid for performing an actual task in the lab.

Real-effort designs achieve less abstraction by trading off experimental control over the participants’ effort costs. However, this lack of control of subjects’ costs restricts the types of tasks that can be used to study a response to incentives. For example, take a simple decision-theoretic model of a real-effort task. In choosing her effort  $e$ , participant  $i$  solves the following problem:

$$e_i^*(w) = \arg \max_{e \in [0, E_i]} w \cdot f_i(e) - c_i(e),$$

where  $w$  denotes a piece rate payment,  $c_i(e)$  denotes the (differentiable) cost of effort she brings into the lab,  $f_i(e)$  denotes her individual production function, and  $E_i$  represents the maximum effort level she can choose to exert (due to physical/time constraints). If we are to use a real-effort task to study the response to incentives in the laboratory, then it must be—at least for the set of incentives considered—that the cost function generates both (i) an interior solution, and (ii) an optimal rule  $e_i^*(w)$  such that that  $f_i(e_i^*(w))$  is not a constant over the set of incentives studied. For example, a task that individuals see as enjoyable will not satisfy these restrictions. If it is satisfying to perform the task, then the marginal cost is negative,  $0 > c'_i(e)$ , and the same optimal effort  $e^*(w) = E_i$  will be chosen for all piece rates  $w \geq 0$ .

---

<sup>1</sup>Reprinted with permission from Springer Nature: Springer, Journal of Economic Science Association, The Slider Task: An Example of Restricted Inference on Incentive Effects, Felipe A. Araujo, Erin Carbone, Lynn Conell-Price, Marli W. Dunietz, Ania Jaroszewicz, Diego Lamé, Rachel Landsman, Lise Vesterlund, Stephanie W. Wang, and Alistair J. Wilson, Copyright 2016

Alternatively, even if the choice of effort is interior, the optimal effort rule could be such that  $f_i(e_i^*(w)) = \hat{y}$  for some constant output  $\hat{y}$ , for any wage  $w$  in the set of incentives studied. This could occur both if the cost of exerting effort is unboundedly large beyond some point or if, perhaps due to low skill, the production function is not very sensitive to changes in effort. In both cases, subjects' behavior will not respond to incentives.

In real-effort tasks the experimenter does not observe effort *per se*, but only the resulting output. In our study, as in most of the literature on real-effort, we use output as a proxy for effort. One inherent difficulty with this approach is that output depends on both effort and ability. For a given increase in incentives, a low-ability subject might substantially increase effort, but not output, while a high-ability subject might only slightly increase effort, and achieve a substantially larger output. Even though our experimental design does not allow for a clean identification of the ability and effort channels, the main contribution of the study, namely testing for a response in incentives, does not hinge in identifying the exact mechanism at work.

The experimental community has been quick to develop creative real effort tasks. In considering easily implementable tasks that are short enough to be run repeatedly, the “slider task” has stood out as being sensitive to incentives. Gill and Prowse (2012, hereafter abbreviated to G&P) introduce the slider task in a study on disappointment aversion. Participants are shown a screen with 48 sliders, where each slider has a range of positions from 0 to 100. Sliders are solved by using the computer's mouse to move the slider's marker (initially placed at 0) to the midpoint of 50. Participants are given two minutes to solve as many sliders as possible, with the participant's chosen effort given by the number of sliders correctly positioned at 50 by the end of the two minutes. The task is normally repeated ten times and cumulative earnings across the entire experiment are given by  $\sum_{t=1}^{10} w_t \cdot f_i(e_i^*(w_t))$ .

Initial evidence from within-subject designs, that the slope of  $f_i(e_i^*(w_t))$  was positive and large, has led to the slider task being used frequently in papers measuring the incentive effects associated with various mechanisms and work environments. However, in contrast to the sensitivity to monetary incentives uncovered in the initial G&P study, more-recent slider-task studies (in particular, those using between-subject designs) find modest or non-existent treatment effects. Our study's main result, obtained from a simple between-subject experimental design, suggest that these recent papers might not be viewed as true null results, but stemming instead from performance in the slider task not being particularly sensitive to changes in the offered incentives.

Where other studies have varied more complex elements of the payoff environment (strategic elements within a game, the nature of feedback, the frame, etc.) ours is a simple between-subject design, focused only on assessing whether the slider task responds to monetary incentives. In fact, we are the only study to look at the slider task as a decision problem, with direct monetary incentives varied between subjects so that experimenter-demand effects can not drive the response to incentives. Building on G&P's implementation of the slider task we conduct three treatments where we vary the piece-rate payment  $w$  that participants receive for each correctly positioned slider: a half cent at the low end, an intermediate two cent treatment, and eight cents at the high end. This sixteen-fold increase in the piece rate corresponds to dramatic differences in participants' potential earnings, with maximum possible performance payments

of \$2.40, \$9.60 and \$38.40, respectively. However, despite substantial differences in the incentives offered, we uncover limited differences in average output: in order of increasing piece rates, we find that subjects complete 26.1, 26.6, and 27.3 sliders per two-minute round. This less than 5 percent increase in response to a 1,500 percent increase in incentives is limited both in magnitude, and relative to the rate of learning and to the individual heterogeneity in ability.<sup>2</sup>

As a real-effort task, the slider task has many attractive characteristics. However, our study shows that the task is not well-suited for uncovering responses to incentives in between-subject designs.

### 3.2 EXPERIMENTAL DESIGN

Our experiments were conducted at the Pittsburgh Experimental Economics Laboratory, using subjects recruited from the student population, randomly assigned to one of three possible treatments.<sup>3</sup> Using a between-subject design the piece rate is held constant throughout an experimental session, so that each subject  $i$  receives a fixed payment per slider of  $w_i \in \{0.5\text{¢}, 2.0\text{¢}, 8.0\text{¢}\}$ .<sup>4</sup> After instructions on the nature of the task, each session began with a two-minute practice round for subjects to become familiar with the slider task. This was followed by ten paying rounds, each of which lasted two minutes. In each round, subjects saw a single screen displaying 48 sliders of equal length and offset from one another, as per G&P.<sup>5</sup> At the end of each round there was a ten second break during which subjects were reminded of how many rounds they had completed, the number of sliders completed, ( $\text{Output}_{it}$ ) and their corresponding earnings from that round ( $w_i \cdot \text{Output}_{it}$ ).<sup>6</sup>

Once the ten paying rounds had concluded, subjects were asked to complete a survey.<sup>7</sup> Only after completing the survey were respondents informed of their total earnings for the session. Subjects were then privately paid, receiving a \$10 participation payment on top of their earnings across the ten rounds  $W_i = \sum_{t=1}^{10} (w_i \cdot \text{Output}_{it})$ .<sup>8</sup>

In order to measure the extent to which the slider task responds to incentives, our study's design adheres closely to that employed in G&P. There are four main differences: i) The G&P design is within subject, where ours is between subject. ii) G&P examine a game between two randomly matched subjects competing over a variable prize; ours examines a decision problem, removing any externalities over payment. iii) The marginal incentives in G&P work

<sup>2</sup>Individual-level heterogeneity can be thought of here as the expected variation across subjects maximal effort  $E_{it}$ , while learning effects can be thought of as within-subject shifts in this variable across time  $t$ .

<sup>3</sup>For consistency, one single member of the project read the instructions for all experimental sessions, and was assisted by another fixed experimenter. All data was collected over the course of two weeks in April of 2015, where all treatments were gender-balanced and interspersed across the data collection period. Initially, a total of three sessions were planned for each treatment; however, a computer error led to subjects' terminals freezing in one round in one session. Another session was therefore added to have three complete sessions for each treatment.

<sup>4</sup>The effective marginal incentives for a risk-neutral subject in G&P varied within a session between 0.15¢ and 6.2¢ with an average of 3.1¢.

<sup>5</sup>The experiment was programmed in z-Tree (Fischbacher, 2007b), and the program KeyTweak was used to disable all arrow keys on the keyboard, thereby ensuring that subjects only used the mouse to complete the slider tasks.

<sup>6</sup>Another difference between our design and the G&P design is that in their experiment subjects were given 2 minute breaks while they waited for their opponent to complete the task.

<sup>7</sup>Data from the survey available from the authors by request.

<sup>8</sup>Subjects in our 0.5¢ treatment had their final payoff  $W_i$  rounded up to the nearest whole cent.



Table 3.1: Summary Statistics

Treatment	Output			N	Hourly Rate
	Avg.	Min	Max		
0.5¢	26.1	6	44	42	\$3.92
2¢	26.6	12	41	43	\$15.95
8¢ <sup>a</sup>	27.3	10	46	63	\$65.46
	(27.4)	(14)	(46)	(45)	(\$65.68)
<b>Total</b>	26.7	10	46	148	

through a probability of winning a prize, where each additional slider completed leads to a one percent increase in the probability of winning a prize; in our experiment the marginal incentives work through a fixed piece rate per slider completed. iv) In G&P peer effects may be present, as subjects observe the other player’s output at the end of each round; in our study there is no feedback on others’ effort levels.

### 3.3 RESULTS

Our experimental results are provided in Tables 3.1 and 3.2. The first table, Table 3.1, reports the average number of sliders completed per round (output), the minimum and maximum output, the total number of subjects  $N$ , and the effective average hourly wage rate (as the incentivized part lasts 20 minutes, this is simply  $3 \cdot W_i$ ). On average, subjects across all of our experiments complete 26.7 sliders in each two-minute period. The lowest number of sliders solved by a subject in any round is ten, where the highest is 46 (two away from the 48 possible). Across treatments, we see that output increases with the piece rate: the average output is 26.1 for the lowest incentive of 0.5¢, somewhat higher at 26.6 for the middle incentive, and at its highest of 27.3 for the 8¢ incentive.

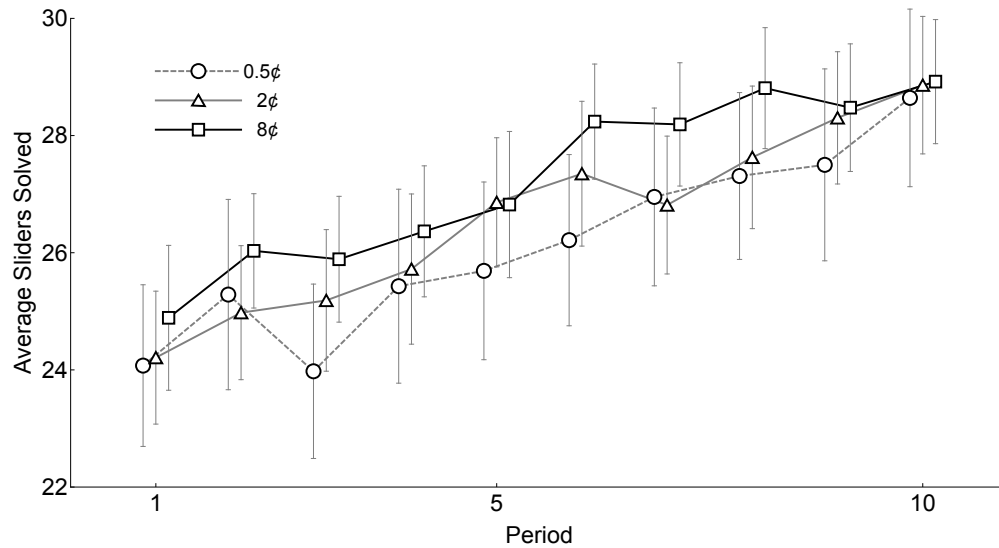
Just from the averages in Table 3.1 it is apparent that the size of the incentive effect is small: going from a piece-rate of 0.5¢ to 2¢ leads to a 0.5 slider increase, and from 2¢ to 8¢ yields a 0.7 slider increase. Though the range of our incentives represents a substantial increase—from an effective hourly rate of about half the US federal minimum to just over \$65 an hour<sup>9</sup>—this 1,500 percent increase in incentives yields less than a 5 percent increase in performance.

Across treatments and sessions, we observe substantial learning. Figure 3.1 presents the round averages for each

<sup>9</sup>By way of comparison, the average lawyer makes an hourly wage of \$64.17 according to the Bureau of Labor Statistics, while the average financial manager makes \$62.61.

of three treatments (where we have additionally provided bars indicating 95 percent confidence intervals, given subject variation). In round one, the average output is 24.2 in both the 0.5¢ and 2¢ treatments, and 24.9 in the 8¢ treatment, though the variation across subjects is large. Across the session, output mostly increases, so that the final output levels in round ten are 28.6 in the 0.5¢ treatments and 28.9 in both the 2¢ and 8¢ treatments. While the output in each treatment appears ordered according to incentives, it is noteworthy that the incentive order is only fully observed in six of the ten rounds.<sup>10</sup>

Figure 3.1: Output Across Rounds



<sup>10</sup>Output in the high wage treatment does appear to flatten out in the last few rounds more so than in the other two treatments. We can think of several reasons why this may be occurring. One possibility, for example, is that it could be the case that higher wages facilitate faster learning. Another possibility is that this trend is due to random chance. Given that our experimental does not let us identify the cause of this trend, this may be an area worth further study by future researchers.

To quantify the effects from incentives while controlling for learning and subject-level variation, we run the following regression:

$$Output_{it} = \beta \cdot \left( \frac{w_i - 0.5}{8 - 0.5} \right) + \sum_{s=2}^{10} \delta_s \cdot 1_{s=t} + \eta + u_i + \epsilon_{it}$$

where  $u_i$  is a subject-level random-effect, and  $\epsilon_{it}$  an idiosyncratic error. The regressions include the treatment as a right-hand-side variable, rescaling the marginal incentive to run *linearly* from zero to one (0.5¢ at the low end, 8¢ at the high, with the 2¢ marginal taking the intermediate value 0.2), and additionally adds nine period dummies as regressors,  $\{\delta_t\}_{t=2}^{10}$  and a constant  $\eta$ . The first column in Table 3.2 reports the estimates for the incentive effect  $\hat{\beta}$ , the initial output level  $\hat{\eta}$  at the beginning of the session, and the average amount of learning across the sessions  $\hat{\delta}_{10}$ . In addition, the table estimates the between-subject standard deviation,  $\hat{\sigma}_u$ , as 3.5 sliders; while the within-subject standard deviation,  $\hat{\sigma}_\epsilon$ , is estimated to be 2.8 sliders.

Unsurprisingly, given the overall averages in Table 3.1, the estimated value of  $\beta$ —where the coefficient represents the estimated marginal effect on sliders solved when moving from the 0.5¢ environment to the 8¢ environment—is close to one slider. Controlling for variation between and within subjects, as well as the across-session learning, the response to incentives is only marginally significant.<sup>11</sup> Interestingly, even our participants appear to be aware that their performance is not motivated by the payment they received. On the survey at the end of the experiment, we find that three-quarters of the participants do not think that there is any lower piece-rate payment at which they would decrease their performance.

Despite a sixteen-fold increase in the piece-rate, the variability in outcomes attributable to incentives is small relative to other variations within the task. In terms of heterogeneity in natural ability, a one slider increase represents under a third of a between-subject standard deviation. In terms of idiosyncratic variation, it represents slightly over a third of a standard deviation. Across the entire session subjects seem to learn to complete more than four additional sliders, relative to their output in round one. So the observed incentive effect represents less than a quarter of the average learning effect.<sup>12</sup>

The second column modifies the regression to use logarithms of the main dependent variable (log of completed sliders) and shifts the right-hand-side incentive variable to measure it in logs.<sup>13</sup> The interpretation of the  $\beta$  estimate in the log regressions is the percentage increase in output as we increase the incentives by 1,500 percent. Though closer to standard significance levels ( $p = 0.066$ ) the estimate of the incentive effect remains very low. Similar to the linear regressions, the 5 percent estimate of the incentive effect is low relative to the 17 percent increase attributable to learning, and to the 12 to 13 percent effect from a within- or between-subject standard deviations.

<sup>11</sup>Including attempted sliders in place of completed sliders, we find an incentive effect of 0.82 sliders ( $p = 0.201$ ).

<sup>12</sup>Allowing rounds to enter into our estimating equation linearly and including a round by treatment interaction term, we fail to reject the null of no differences in learning effects between treatments ( $p = 0.64$ ).

<sup>13</sup>More exactly, the RHS incentive variable in our log regressions is rescaled and renormalized so that the incentive runs linearly from zero to one with the 2¢ marginal incentive taking the value of 0.5 (as our wage rates are  $2^{-1}$ ,  $2^1$  and  $2^3$ ), where our linear regression had 2¢ representing just 20 percent of the overall shift in incentives.

Table 3.2: Random-Effect Regressions

Estimate	Our Data		G&P		G&P Restricted	
	Linear	Log	Linear	Log	Linear	Log
Incentive Effect, $\beta$	1.05 <sup>a</sup> (0.65)	0.05 <sup>a</sup> (0.03)	3.27 <sup>c</sup> (0.75)	0.12 (0.04)	2.67 <sup>c</sup> (0.65)	0.08 (0.02)
Initial Output, $\hat{\eta}$	23.98 (0.48)	3.15 (0.02)	21.11 (0.89)	2.95 (0.06)	22.70 (0.72)	3.10 (0.03)
Learning Effect, $\hat{\delta}_{10}$	4.34 (0.32)	0.17 (0.01)	4.35 (0.71)	0.19 (0.05)	4.24 (0.62)	0.16 (0.02)
Between Std. Dev., $\hat{\sigma}_u$ <sup>b</sup>	3.47 (0.29)	0.13 (0.01)	5.40 (0.68)	0.27 (0.06)	3.91 (0.33)	0.15 (0.01)
Within Std. Dev., $\hat{\sigma}_\epsilon$ <sup>b</sup>	2.77 (0.12)	0.12 (0.11)	3.87 (0.47)	0.29 (0.08)	3.22 (0.32)	0.13 (0.02)

Numbers in parentheses are standard errors.

*a*- A programming error led to one round in a single 8¢ marginal session hanging, so that data was not recorded for that round. Results are similar if we remove the entire session from our analysis and so we conduct our analysis on the entire data excluding this one round. The estimated incentive effect in Table 3.1(B) when dropping this session is 1.16 sliders, where the test for significance has a *p*-value of 0.107.

*b*- Standard errors for between and within standard deviations are drawn from a bootstrap of size 1,000 that resamples across subjects, then subject-rounds.

*c* -Marginal incentives for G&P first movers are calculated relative to our upper and lower incentive treatments. Because of this  $\hat{\eta}$  has the interpretation of average output (average log of output) in round one at a 0.5¢ incentive in all regressions, and  $\beta$  has the interpretation as the estimated marginal effect of going from a 0.5¢ environment to an 8¢ environment in all regressions.

*d* -In G&P Restricted we excluded all of their participants whose performance was lower than our worst performing subject.

Table 3.3: Power Calculations

	<b>\$0.005 to \$0.02</b>			<b>\$0.02 to \$0.08</b>			<b>\$0.005 to \$0.08</b>		
	$p=0.1$	$p=0.05$	$p=0.01$	$p=0.1$	$p=0.05$	$p=0.01$	$p=0.1$	$p=0.05$	$p=0.01$
$n=20$	0.150	0.083	0.020	0.238	0.150	0.047	0.383	0.270	0.104
$n=30$	0.170	0.099	0.026	0.300	0.199	0.070	0.496	0.373	0.171
$n=40$	0.191	0.114	0.033	0.360	0.249	0.097	0.597	0.472	0.243
$n=50$	0.212	0.130	0.039	0.418	0.300	0.126	0.681	0.561	0.320
$n=60$	0.234	0.146	0.045	0.471	0.349	0.156	0.748	0.639	0.397
$n=70$	0.254	0.163	0.054	0.520	0.395	0.190	0.802	0.706	0.470
$n=80$	0.275	0.178	0.062	0.569	0.442	0.223	0.847	0.763	0.540
$n=90$	0.294	0.194	0.069	0.610	0.486	0.257	0.882	0.809	0.604
$n=100$	0.314	0.208	0.078	0.651	0.528	0.293	0.910	0.848	0.661
$n=150$	0.406	0.290	0.121	0.801	0.702	0.467	0.978	0.955	0.861
$n=200$	0.491	0.367	0.170	0.891	0.821	0.618	0.995	0.988	0.950

Even if we disregard the small magnitude of the effect and only focus on significance, the slider task is severely underpowered for uncovering a response to incentives with a typical experimental sample size. Table 3.3 shows the power calculations for all pairwise treatment comparisons; 90% power is unattainable with fewer than 200 subjects per treatment for both of our fourfold increases in incentives even at a 10% significance, while for a 1500% increase in incentives we would need over 100 subjects per treatment to detect a significant difference 90% of the time at a 5% significance level.

### 3.4 DISCUSSION

With a supply elasticity of only 0.003, our between-subject design finds that effort exerted on the slider task, proxied by output, is very inelastic. We now examine how our results compare to G&P.

In G&P, two players  $i$  (a first mover) and  $j$  (a second mover) are randomly matched and compete to win a common prize of size  $100 \cdot w_{it}$  cents, drawn randomly from an interval. The probability of player  $i$  winning the prize is given by  $\frac{1}{100} (50 + \text{Output}_{it} - \text{Output}_{jt})$ , so for a risk-neutral participant the expected marginal incentive is  $w_{it}$ .<sup>14</sup> The sequencing of the game is such that the first mover's output ( $\text{Output}_{it}$ ) is observed by the second mover  $j$ , and the second mover's response is the main focus in G&P. In looking at the response to incentives, we follow Gill and Prowse (2015) and look only at the first movers.

As noted earlier, the first mover's task in G&P is different from that in our study: i) Their sessions have within-subject variation over the incentive  $w_{it}$ , so subjects have some knowledge of the experimenter's question, thus demand effects may be present; ii) the tournament structure has own output inflicting a negative externality on the other player; iii) payment is incentivized only probabilistically; and iv) there is feedback on other participants' output levels. Changes in levels may plausibly come from any of these differences, and future research might help isolate each of these channels. However, it is still of interest to compare the magnitudes of the incentive effect in G&P and our study.

Paralleling the regression results from our data in the first pair of columns in Table 3.2, the next two pairs of columns provide similar random-effects regressions from the G&P data. The first pair of G&P regressions provide results under the linear and log specification for the  $N = 60$  first-movers.<sup>15</sup> The coefficient  $\tilde{\beta}$  reflects the estimate from the G&P data for the incentive effect in our experiment, showing that the G&P data predicts a significant 3.26 sliders increase as the marginal incentive is raised from 0.5¢ to 8¢. Our incentive estimate  $\hat{\beta}$  from Table 3.2 is much smaller and is significantly different from the G&P estimate ( $p = 0.000$ ).

The high incentive effect stems in part from a number of first-mover subjects who have very low output levels in the G&P data. There could be several reasons for producing low output. One possibility that exists in G&P but not in

<sup>14</sup>The raw prizes in G&P are drawn uniformly over  $\{\pounds 0.10, \pounds 0.20, \dots, \pounds 3.90\}$ . We transform these to expected marginal incentives for a risk-neutral agent, and then convert to US cents at a conversion rate of  $\pounds 0.65 = 100\text{¢}$ .

<sup>15</sup>To distinguish between estimates on our data and G&P's we will use the notation  $\hat{\beta}$ ,  $\hat{\eta}$ , etc., for estimates from our data, and  $\tilde{\beta}$ ,  $\tilde{\eta}$ , etc., for estimates from the G&P data.

our study is that subjects might be trying to pick the efficient outcome (both exerting zero effort and equally splitting the chance to win the prize).<sup>16</sup> As a partial control for this, we re-run the same random-effects regressions excluding the G&P first-movers whose average output across the ten rounds is *lower than the lowest subject average* in our between-subject data (18.5 sliders, from the 0.5¢ treatment). This excludes six subjects, representing ten percent of the G&P first mover subjects.<sup>17</sup>

The regression results for the G&P subsample are given in the final pair of columns in Table 3.2. Though the estimated incentive effect is lower than the full sample—decreasing to 2.67 sliders—our estimate is still significantly different ( $p = 0.012$ ). Moreover, despite the large differences in the estimated incentive effects, the other regression coefficients are remarkably similar.

Looking at the results in the linear specification with  $N = 54$  (where we remove subjects in the left tail of the distribution), and comparing them to our results in the first column in Table 3.2, we find many commonalities. First, subjects on average increase performance across the session by approximately four sliders ( $\hat{\delta}_{10}$  and  $\tilde{\delta}_{10}$  are not significantly different).<sup>18</sup> Second, though the initial output level estimates of  $\eta$  are significantly higher in our sessions at 24 sliders in comparison to 22.7 in G&P, the size of the difference is quantitatively small.<sup>19</sup> Third, between- and within-subject standard deviations for output after controlling for the incentive effects ( $\sigma_u$  and  $\sigma_\epsilon$ , respectively) are very similar, though in both cases the estimated variation in our experiments is smaller than in G&P.

Comparing our results to those of G&P, it is hard not to attribute the majority of the observed incentive effect to some combination of a within-subject effect (demand or peer effects) and a strategic or social effect (with the negative externality pushing subjects to exert low effort). While we leave it to future research to disentangle which of these factors are driving the additional incentive effects, it is clear that the effect observed in our data can at best be described as marginal.

### 3.5 CONCLUSION

Using a between-subject design, we examine how performance on the slider task responds to changes in monetary incentives. Despite a 1,500 percent increase in incentives we find only a five percent increase in output. With a supply-elasticity of only 0.003 our results show that the slider task is poorly suited for studying the response to incentives in between-subject designs.

Three recent studies point to techniques which might offer more-constructive results for real-effort tasks in the lab. Gächter et al. (2015) introduce a ball-catching task where the cost of effort is directly manipulated by the experimenter.

<sup>16</sup>Gill & Prowse (2015) note that 2 subjects (whom we will also exclude) appear to have difficulty positioning sliders at exactly 50 until a few rounds into the session.

<sup>17</sup>Note that only subjects with low *average* performance are eliminated from the data. Data from subjects with particular rounds with less than 19 sliders completed are still included in the analysis, provided that the subject's average across the session is above 18.5 sliders.

<sup>18</sup>All three of our treatments, as well as both movers in G&P show fairly consistent increases in average output across the session.

<sup>19</sup>A joint regression across both sets of data indicates no significant difference over the two constants ( $p = 0.123$ ).

With suitable parameterizations, interior solutions can therefore be ensured. Less directly, [Corgnet et al. \(2014\)](#) and [Eckartz \(2014\)](#) examine a variety of real-effort tasks and find that the presence of outside leisure activities and paid outside options, respectively, lead to stronger incentive effects. These different approaches—the one with greater experimental control, the other with greater flexibility extended to subjects—suggest possible solutions for researchers wishing to use the slider task in a between-subject design.

While there are several reasons that the incentive effect might be larger in the G&P data, our study motivates future research on the potential greater sensitivity in within-subject designs.<sup>20</sup> One explanation for stronger results in within-subject designs is that they allow for better controls for the large variation in individual-level ability of the slider task.<sup>21</sup> An alternative, but undesirable explanation, is that the additional response is an experimenter-demand effect. Future research is needed to identify the cause of these between/within-subject differences.

Whatever the cause, a reasonable criterion when using *any* real-effort task to study the incentives is a demonstrated response to explicit monetary incentives between subject. Statistical significance aside, desirable tasks should be able to demonstrate an incentive effect which is measurable in magnitude, and large relative to uncontrolled variation within the task (individual ability, learning, *etc.*). With respect to this above criterion, our study sounds a cautious note for the slider task. While the task has many appealing properties it is underpowered for uncovering a response to incentives.

---

<sup>20</sup>In mirroring responses to incentives in labor markets one may wish to think of within designs as capturing short-term effects and between designs as capturing long-run responses.

<sup>21</sup>If this is the main channel, one way to reduce noise from individual heterogeneity is to measure baseline ability via a common task with a fixed incentive level at the start of each treatment, à la [Lilley and Slonim \(2014\)](#), with subsequent tasks chosen with the desired between-subject variation.



## APPENDIX

### ADDITIONAL FIGURES

Figure A.1: Difficulty Adjusted Compensation - First Four Periods

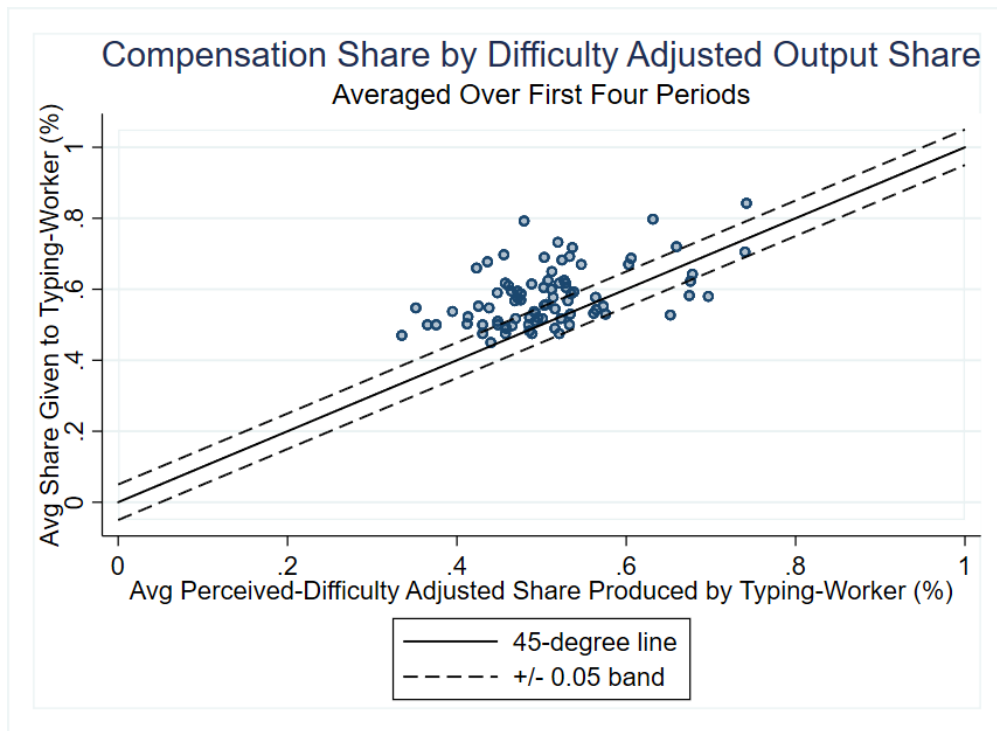
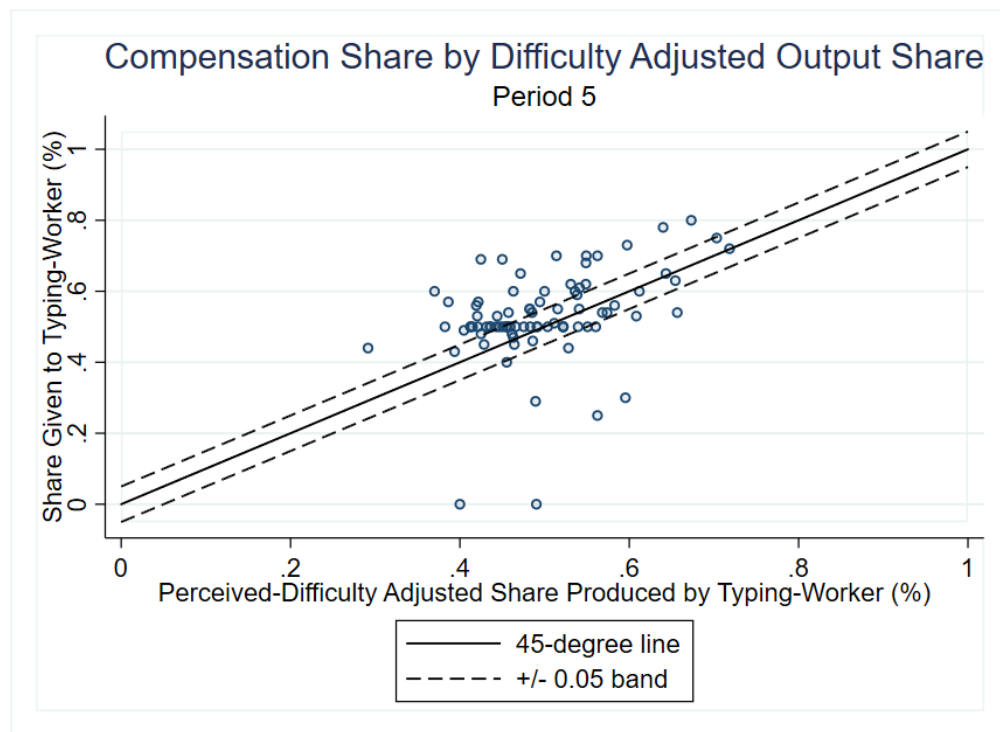


Figure A.2: Difficulty Adjusted Compensation - All Five Periods



## BIBLIOGRAPHY

- Aaberge, R. (1997). Interpretation of changes in rank-dependent measures of inequality. *Economics Letters*, 55(2):215–219.
- Adams, S. M., Gupta, A., and Leeth, J. D. (2009). Are female executives over-represented in precarious leadership positions? *British Journal of Management*, 20(1):1–12.
- Albanesi, S., Olivetti, C., and Prados, M. J. (2015). Gender and dynamic agency: Theory and evidence on the compensation of top executives. *Working Paper*.
- Araujo, F. A., Carbone, E., Conell-Price, L., Dunietz, M. W., Jaroszewicz, A., Landsman, R., Lamé, D., Vesterlund, L., Wang, S. W., and Wilson, A. J. (2016). The slider task: an example of restricted inference on incentive effects. *Journal of the Economic Science Association*, 2(1):1–12.
- Babcock, L. and Laschever, S. (2003). *Women Don't Ask: Negotiation and the Gender Divide*.
- Bailey, M. J. and DiPrete, T. A. (2016). HHS Public Access. *The Russell Sage Foundation Journal of the Social Sciences : RSF*, 2(4):1–32.
- Becker-Blease, J. R., Elkinawy, S., Hoag, C., and Stater, M. (2016). The Effects of Executive, Firm, and Board Characteristics on Executive Exit. *Financial Review*, 51(4):527–557.
- Becker-Blease, J. R., Elkinawy, S., and Stater, M. (2010). The Impact of Gender on Voluntary and Involuntary Departure. *Economic Inquiry*, 48(4):1102–1118.
- Bertrand, M. and Mullainathan, S. (2001). Are CEOs Rewarded for Luck? The Ones Without Principals Are. *The Quarterly Journal of Economics*, 116(3):901–932.
- Bidwell, M. (2011). Paying More to Get Less: The Effects of External Hiring versus Internal Mobility. *Administrative Science Quarterly*, 56(3):369–407.
- Blau, F. D. and Kahn, L. M. (2016). The Gender Wage Gap: Extent, Trends, and Explanations. *NBER Working Paper Series*, 21913:1–75.
- Bolton, G. and Werner, P. (2016). The influence of potential on wages and effort. *Experimental Economics*, 19(3):535–561.
- Bowles, H. R., Babcock, L., and Lai, L. (2007). Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and Human Decision Processes*, 103(1):84–103.
- Bowles, H. R., Babcock, L., and McGinn, K. L. (2005). Constraints and triggers: Situational mechanics of gender in negotiation. *Journal of Personality and Social Psychology*, 89(6):951–965.

- Brett, J. M. and Stroh, L. K. (1997). Jumping Ship: Who Benefits From an External Labor Market Career Strategy? *Journal of Applied Psychology*, 82(3):331–341.
- Bull, C., Schotter, A., and Weigelt, K. (1987). Tournaments and piece rates: An experimental study. *Journal of Political Economy*, pages 1–33.
- Cameron, K. S., Freeman, S. J., and Mishra, A. K. (1993). Downsizing and redesigning organizations.
- Castanias, R. P. and Helfat, C. E. (1991). Managerial Resources and Rents. *Journal of Management*, 17(1):155–171.
- Cook, A. and Glass, C. (2014). Women and Top Leadership Positions: Towards an Institutional Analysis. *Gender, Work and Organization*, 21(1):91–103.
- Cooper, D. J. and Kühn, K. U. (2014). Communication, renegotiation, and the scope for collusion. *American Economic Journal: Microeconomics*, 6(2):247–278.
- Corgnet, B., Hernán-González, R., and Schniter, E. (2014). Why real leisure really matters: Incentive effects on real effort in the laboratory. *Experimental Economics*, pages 1–18.
- Deaux, K. (1979). Self-evaluations of male and female managers. *Sex Roles*, 5(5):571–580.
- Deaux, K. and Emswiller, T. (1974). Explanations of successful performance on sex-linked tasks: What is skill for the male is luck for the female. *Journal of Personality and Social Psychology*, 29(1):80–85.
- DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, 47(2):315–372.
- Dikolli, S. S., Mayew, W. J., and Nanda, D. (2014). CEO tenure and the performance-turnover relation. *Review of Accounting Studies*, 19(1):281–327.
- Dittrich, M., Knabe, A., and Leipold, K. (2014). Gender Differences in Experimental Wage Negotiations. *Economic Inquiry*, 52(2):862–873.
- Dreher, G. F., Lee, J.-Y., and Clerkin, T. A. (2011). Mobility and Cash Compensation: The Moderating Effects of Gender, Race, and Executive Search Firms. *Journal of Management*, 37(3):651–681.
- Eckartz, K. (2014). Task enjoyment and opportunity costs in the lab: The effect of financial incentives on performance in real effort tasks. Jena Economic Research Papers 2014-005.
- Egan, M. L., Matvos, G., and Seru, A. (2017). When Harry Fired Sally: The Double Standard in Punishing Misconduct.
- Eil, D. and Rao, J. M. (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2):114–138.
- Ertac, S. (2011). Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior and Organization*, 80(3):532–545.
- Ewing Marion Kauffman Foundation (2017). 2017 Kauffman Index of Startup Activity Ewing Marion Kauffman Foundation. Technical report.
- Exley, C., Niederle, M., and Vesterlund, L. (2018). Knowing When to Ask: The Cost of Leaning-in. *Working Paper*.
- Fama, E. F. and French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, 43(2):153–193.

- Fee, C. E., Hadlock, C. J., Huang, J., and Pierce, J. R. (2015). Robust Models of CEO Turnover: New Evidence on Relative Performance Evaluation.
- Fehr, E., Kirchsteiger, G., and Riedl, A. (1993). Does fairness prevent market clearing? an experimental investigation. *Quarterly Journal of Economics*, pages 437–459.
- Fernandez-Mateo, I. and Fernandez, R. M. (2016). Bending the Pipeline? Executive Search and Gender Inequality in Hiring for Top Management Jobs. *Management Science*, 62(12):3636–3655.
- Finkelstein, S., Hambrick, D., and Cannella, A. (2009). *Strategic leadership: Theory and research on executives, top management teams, and boards*.
- Fischbacher, U. (2007a). z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics*, 10(2):171–178.
- Fischbacher, U. (2007b). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178.
- Gächter, S., Huang, L., and Sefton, M. (2015). Combining “real effort” with induced effort costs: The ball-catching task. IZA Discussion Paper 9041.
- Gayle, G.-L., Golan, L., and Miller, R. A. (2012). Gender Differences in Executive Compensation and Job Mobility. *Journal of Labor Economics*, 30(4):829–872.
- Gill, D. and Prowse, V. (2012). A Structural Analysis of Disappointment Aversion in a Real Effort Competition. *American Economic Review*, 102(1021):469–503.
- Gross, T., Guo, C., and Charness, G. (2015). Merit pay and wage compression with productivity differences and uncertainty. *Journal of Economic Behavior and Organization*, 117:233–247.
- Guay, W., Taylor, D. J., and Xiao, J. J. (2014). Adapt or Perish: Evidence of CEO Adaptability to Industry Shocks.
- Guest, P. M. (2016). Executive Mobility and Minority Status. *Industrial Relations*, 55(4):604–631.
- Guvenen, F., Kaplan, G., and Song, J. (2014). The Glass Ceiling and the Paper Floor: Gender Differences among Top Earners, 1981–2012. *National Bureau of Economic Research Working Paper Series*, pages 1981–2012.
- Hambrick, D. C. and Mason, P. A. (1984). Upper Echelons: The Organization as a Reflection of Its Top Managers. *Academy of Management Review*, 9(2):193–206.
- Haslam, S. A. and Ryan, M. K. (2008). The road to the glass cliff: Differences in the perceived suitability of men and women for leadership positions in succeeding and failing organizations. *Leadership Quarterly*, 19(5):530–546.
- Hennessey, S. M., MacDonald, K., and Carroll, W. (2014). Is there a “glass cliff or a solid ledge for female appointees To the Board of Directors? *Journal of Organizational Culture, Communications & Conflict*, 18(2):125–139.
- Holmström, B. (1982). Moral hazard in teams. *Bell Journal of Economics*, 11(2):74–91.
- Homroy, S. (2015). Are CEOs replaced for poor performance? Effects of takeovers and governance on CEO turnover. *Scottish Journal of Political Economy*, 62(2):149–170.
- Jensen, M. C. and Murphy, K. J. (1990). Performance pay and top-management incentives. *Journal of political economy*, 98(2):225–264.
- Jenter, D. and Kanaan, F. (2006). CEO Turnover and Relative Performance Evaluation.

- Jenter, D. and Kanaan, F. (2015). CEO Turnover and Relative Performance Evaluation. *The Journal of Finance*, 70(5):2155–2184.
- Jenter, D. and Lewellen, K. (2010). Performance-induced CEO turnover.
- Jones, L. E. and Schoonbroodt, A. (2016). Baby busts and baby booms: The fertility response to shocks in dynastic models. *Review of Economic Dynamics*, 22:157–178.
- Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–292.
- Kark, R. and Eagly, A. H. (2010). Gender and Leadership: Negotiating the Labyrinth. In *Handbook of Gender Research in Psychology*, pages 443–468.
- Koszegi, B. and Rabin, M. (2006). A Model of Reference-Dependent Preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.
- Kray, L. (2015). The best way to eliminate the gender pay gap? Ban salary negotiations.
- Krishnan, H. A. and Park, D. (1998). Effects of Top Management Team Change on Performance in Downsized US Companies. *Management International Review*, 38(4):303–319.
- Kulich, C., Lorenzi-Cioldi, F., Iacoviello, V., Faniko, K., and Ryan, M. K. (2015). Signaling change during a crisis: Refining conditions for the glass cliff. *Journal of Experimental Social Psychology*, 61:96–103.
- Landis, J. R. and Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159.
- Liebersohn, C. J. (2016). The Cyclicity of Executive Turnover.
- Lilley, A. and Slonim, R. (2014). The price of warm glow. *Journal of Public Economics*, 114:58–74.
- Marzilli Ericson, K. M. and Fuster, A. (2014). The Endowment Effect. *Annual Review of Economics*, 6(1):555–579.
- Mazei, J., Hüffmeier, J., Freund, P. A., Stuhlmacher, A. F., Bilke, L., and Hertel, G. (2015). A meta-analysis on gender differences in negotiation outcomes and their moderators. *Psychological Bulletin*, 141(1):85–104.
- Möbius, M. M., Niederle, M., Niehaus, P., and Rosenblat, T. S. (2014). Managing Self-Confidence. *Working Paper*.
- Mulcahy, M. and Linehan, C. (2014). Females and Precarious Board Positions: Further Evidence of the Glass Cliff. *British Journal of Management*, 25(3):425–438.
- Murphy, K. J. (1999). Chapter 38 Executive compensation. *Handbook of Labor Economics*, 3 PART(2):2485–2563.
- Nalbantian, H. R. and Schotter, A. (1997). Productivity under group incentives: An experimental study. *American Economic Review*, pages 314–341.
- Niederle, M. and Vesterlund, L. (2007). Do Women Shy Away From Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics*, 122(3):1067–1101.
- Oster, E. (2016). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics*, pages 1–18.
- Parrino, R. (1997). CEO turnover and outside succession: A cross-sectional analysis. *Journal of Financial Economics*, 46(2):165–197.

- Quintana-García, C. and Elvira, M. M. (2017). The Effect of the External Labor Market on the Gender Pay Gap among Executives. *ILR Review*, 70(1):132–159.
- Ryan, M. K. and Haslam, S. A. (2005). The glass cliff: Evidence that women are over-represented in precarious leadership positions. *British Journal of Management*, 16(2):81–90.
- Ryan, M. K., Haslam, S. A., Hersby, M. D., and Bongiorno, R. (2011). Think crisis–think female: The glass cliff and contextual variation in the think manager–think male stereotype. *Journal of Applied Psychology*, 96(3):470–484.
- Schaller, J. (2016). Booms, Busts, and Fertility Testing the Becker Model Using Gender-Specific Labor Demand. *Journal of Human Resources*, 51(1):1–29.
- Schneider, D. (2015). The Great Recession, Fertility, and Uncertainty: Evidence From the United States. *Journal of Marriage and Family*, 77(5):1144–1156.
- Schotter, A. and Weigelt, K. (1992). Asymmetric tournaments, equal opportunity laws, and affirmative action: Some experimental results. *Quarterly Journal of Economics*, pages 511–539.
- Selody, K. (2010). Is the Risk Worth the Reward for Top Female Executives?
- Sobotka, T., Skirbekk, V., and Philipov, D. (2011). Economic recession and fertility in the developed world.
- Stefanescu, I., Wang, Y., Xie, K., and Yang, J. (2015). Pay Me Now ( and Later ): Pension Benefit Manipulation before Plan Freezes and Executive Retirement.
- Study, C. S. (2016). CEO Success Study. Technical report.
- The Rockefeller Foundation and Global Strategy Group (2016). Does the media influence how we perceive women in leadership ? CEOs and gender : a media analysis. Technical report.
- Weber, R., Camerer, C., Rottenstreich, Y., and Knez, M. (2001). The illusion of leadership: Misattribution of cause in coordination games. *Organization Science*, 12(5):582–598.
- Yonker, S. E. (2016). Geography and the Market for CEOs. *Management Science*, 63(3):609–630.